

3D Chromatin Remodeling Potentiates Transcriptional Programs Driving Cell Invasion

Benjamin Lebeau

Lady Davis Institute for Medical Research

Maika Jangal

Lady Davis Institute of the Jewish General Hospital

Tiejun Zhao

Lady Davis Institute of the Jewish General Hospital

Cheng Kit Wong

Lady Davis Institute, McGill University

Nolan Wong

Lady Davis Institute

Eduardo Cepeda Canedo

Lady Davis Institute for Medical Research

Steven Hebert

Lady Davis Research Institute <https://orcid.org/0000-0001-6267-8595>

Adriana Aguilar-Mahecha

Jewish General Hospital

Catherine Chabot

Jewish General Hospital

Marguerite Buchanan

Jewish General Hospital

Rachel Catterall

Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montreal, QC, H3A 1A3, Canada

Luke McCaffrey

McGill University <https://orcid.org/0000-0002-1754-8828>

Genevieve Deblois

Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

Claudia Kleiman

Jewish General Hospital <https://orcid.org/0000-0002-5158-7126>

Morag Park

McGill University <https://orcid.org/0000-0001-5400-606X>

Mark Basik

Michael Witcher (✉ michael.witcher@mcgill.ca)

Lady Davis Institute and Segal Cancer Centre, Jewish General Hospital <https://orcid.org/0000-0002-4466-9166>

Article

Keywords: Breast Cancer, CTCF, subTAD, TAD, H3K27ac, Epigenetic, Chromatin Conformation, SNAI1, mTor Inhibitor

Posted Date: October 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-956814/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

1 **Title:**

2 **3D Chromatin Remodeling Potentiates Transcriptional Programs Driving Cell Invasion**

3 **Authors:**

4 Benjamin Lebeau^{1,2}, Maïka Jangal², Tiejun Zhao², Cheng Kit Wong^{1,2}, Nolan Wong^{2,3}, Eduardo
5 Cepeda Cañedo^{1,2}, Steven Hébert², Adriana Aguilar-Mahecha², Catherine Chabot², Marguerite
6 Buchanan², Rachel Catterall^{1,4}, Luke McCaffrey^{1,4}, Geneviève Deblois^{5,6}, Claudia Kleinman^{2,7},
7 Morag Park^{4,8}, Mark Basik^{1,9} and Michael Witcher^{1,2*}

8 1. Department of Experimental Medicine, Faculty of Medicine, McGill University,
9 Montréal, QC H4A 3J1, Canada.

10 2. Lady Davis Research Institute, Jewish General Hospital, Montréal, QC H3T 1E2,
11 Canada.

12 3. Department of Biochemistry, The University of British Columbia, Vancouver, BC V6T
13 1Z4, Canada.

14 4. Rosalind & Morris Goodman Cancer Research Centre, McGill University, Montréal, QC
15 H3A 1A3, Canada.

16 5. Faculty of Pharmacy, Université de Montréal, Montreal, QC H3T 1J4, Canada.

17 6. Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, QC
18 H3T 1J4, Canada.

19 7. Department of Human Genetics, Faculty of Medicine, McGill University, Montréal, QC
20 H4A 3J1, Canada.

21 8. Department of Biochemistry, McGill University, Montréal, QC H4A 3J1, Canada.

22 9. Departments of Surgery and Oncology, Faculty of Medicine, McGill University,
23 Montréal, QC H4A 3J1, Canada.

24

25 * Corresponding Author and Lead Contact. Correspondence: michael.witcher@mcgill.ca

26

Abstract:

The contribution of deregulated chromatin architecture, including topologically associated domains (TADS), to cancer progression remains ambiguous. CTCF is a central regulator of higher-order chromatin structure that undergoes copy number loss in over half of all breast cancers, but the impact of this defect on epigenetic programming and chromatin architecture remains unclear. We find that under physiological conditions, CTCF organizes sub-TADs to limit the expression of oncogenic pathways, including PI3K and cell adhesion networks. Loss of a single CTCF allele potentiates cell invasion through compromised chromatin insulation and a reorganization of chromatin architecture and histone programming that facilitates de novo promoter-enhancer contacts. However, this change in the higher-order chromatin landscape leads to a vulnerability to inhibitors of mTOR. These data support a model whereby sub-TAD reorganization drives both the modification of histones at de novo enhancer promoter-contacts and transcriptional upregulation of oncogenic transcriptional networks.

Keywords: Breast Cancer, CTCF, subTAD, TAD, H3K27ac, Epigenetic, Chromatin Conformation, SNAI1, mTor Inhibitor.

44 **Introduction:**

45

46 Hierarchical nuclear organization of chromatin plays essential roles during development and cell
47 specification^{1,2}. As such, mapping and understanding the functionality of three dimensional (3D)
48 chromatin structure is now at the forefront of epigenetics research. Based on the advent of Hi-C
49 sequencing technology³, we know that the entire genome is organized into an assembly of
50 Topologically Associated Domains (TADs). TADs comprise 100kb to 1Mb regions of chromatin
51 defined as a contiguous region enriched for DNA-DNA contacts between loci within the TAD,
52 with few interactions outside of the TAD⁴. TADs are commonly anchored by CTCF together
53 with the cohesin complex, establishing a stable chromatin domain⁵⁻⁹. Within TADs, smaller
54 regions of self-interaction, called subTADs, add an additional layer of complexity to 3D
55 chromatin architecture¹⁰.

56 TADs and subTADs regulates gene transcription in mechanistically similar ways. By confining
57 chromatin interactions to a particular region, they promote local interactions between cis-
58 regulatory elements, such as enhancer-promoter interactions, while insulating from outside cis-
59 regulatory elements. This allows for the specific pairings of promoters and enhancers required
60 for proper temporal regulation of gene expression¹¹. TAD and subTADs structures also affect the
61 deposition of diverse epigenetic marks. This may be achieved by blocking the spread of histone
62 modifications outside of the domain¹² or alternatively, histone marks may be directly influenced
63 by DNA contacts. For example, the deposition of H3K27ac on active enhancers and promoters is
64 strongly associated with, and maybe dependant upon, DNA-DNA contacts at transcriptionally
65 active regions^{13,14}. Organization of chromatin into subTADs facilitates a more precise and
66 dynamic local regulation of transcription than TADs alone would allow. Indeed, dynamic
67 changes in subTAD organization drive transcriptional events of differentiation and cell identity,
68 while TAD boundaries are mostly stable during these processes¹.

69 Proper TAD/subTAD organization is essential for temporal control of gene expression during
70 development^{15,16}. While aberrant activation of developmental programs appears to play an
71 important role in tumor progression, it is unclear that global reorganization of chromatin domains
72 is involved in this process. There is evidence that altered TAD or subTAD organization locally at
73 oncogenic loci may promote tumor initiation via aberrant changes to gene transcription. For

74 example, in hematological tumors, reorganization of DNA contacts within a TAD encompassing
75 a portion of the HOXA cluster facilitates transcriptional activation, thereby driving
76 leukemogenesis¹⁷. Genome-wide analysis of chromatin contacts using relevant models of tumor
77 initiation and progression are clearly needed to provide further insights into a potential role of
78 TAD reorganization in these processes.

79 CCCTC-binding factor (CTCF) is a multifunctional epigenetic regulatory protein critical for the
80 establishment and maintenance of TAD boundaries¹⁸. The loss of CTCF binding is known to
81 cause the loss of insulation of TAD/subTAD boundaries, resulting in chromatin domain shift or
82 fusion. The degradation of CTCF as also been shown to affect specifically global subTAD
83 organization¹⁹. CTCF binding sites are mutated in a variety of cancers²⁰, but due to the complex,
84 multifunctional nature of CTCF, it is challenging to predict the transcriptional changes that might
85 ensue after the loss of CTCF at specific binding sites. Consistent with this, disruptions of
86 individual boundaries have been reported to both increase or decrease oncogene expression,
87 depending on the context²¹⁻²³.

88 Considering the central role of CTCF in maintenance of genomic TADs, it is not surprising that
89 CTCF knockout leads to lethality at very early stages of embryonic development^{24,25}. Although
90 not lethal, the loss of heterozygosity at the CTCF locus is also detrimental to cellular
91 homeostasis²⁶ and there is evidence that CTCF acts as an haploinsufficient tumor suppressor
92 gene^{27,28}, with its loss potentially impacting hematopoietic tumor initiation²⁷, but the impact on
93 solid tumors remains ambiguous. Consistent with a role as a tumor suppressor, data from the
94 Cancer Genome Atlas reveals that sixty-three percent of all breast tumors harbor CTCF copy
95 number loss (CNL), which correlates with lowered expression²⁹. There is substantial clinical data
96 that 16q22.1 is related to tumor progression³⁰⁻³². However, CNL at 16q22.1 concurrently deletes
97 additional genes adjacent to CTCF, including the tumor suppressor CDH1, confounding our
98 ability to define the impact of CTCF loss. While it has been hypothesized that fluctuations in
99 CTCF levels may impact chromatin looping this has not been formally examined³³. Thus, it
100 remains unclear whether transcriptional networks and topological features may be deregulated in
101 breast epithelium undergoing CTCF copy number loss.

102 In the current study we find that CTCF hemizyosity in mammalian mammary epithelial cells
103 potentiates subTAD reorganization and cell invasion. Likewise, reintroduction of CTCF into

104 patent derived xenograft (PDX) lines with CTCF CNL prohibits their invasion. In these models,
105 restructuring of chromatin architecture, especially at the subTAD level, drives activation of the
106 phosphatidylinositol 3-kinase (PI3K) pathway and overexpression of the classical oncogene
107 SNAI1. These changes are associated with epigenetic reprogramming of H3K27ac and
108 H3K4me3 at regulatory regions. We also pose that these altered transcriptional events predicts
109 sensitivity to mTor inhibitors³⁴, that potently repress the invasive capacity of CTCF hemizygous
110 cells.

111

112 **Results:**

113

114 **CTCF hemizyosity in breast epithelium Promotes Invasiveness and Disorganized** 115 **Growth**

116

117 CTCF single allele deletions are prevalent in a majority of breast tumors²⁹. To better understand
118 the consequences of this genetic aberration we first aimed to generate a model of CTCF copy
119 number loss in human breast epithelium. Using a CRISPR-Cas9 approach, we achieved a
120 knockout of one allele of CTCF in the mammary epithelial cell line MCF10A ([Figure 1a](#)). Here,
121 guide RNAs targeting exon 6 of CTCF led to a deletion resulting in an early stop codon within
122 one allele of CTCF. These clonal-derived, hemizygous knockouts of CTCF led to a reduction of
123 CTCF protein levels by fifty to sixty-six percent ([Figure 1a](#)). These reductions in CTCF tightly
124 mimic the reduced levels of CTCF observed in PDX tumors from triple-negative breast cancer
125 patients ([Supplementary Figure 1a](#)).

126 Using this model, we screened for several classical oncogenic phenotypes including altered
127 morphology, increased proliferation, anchorage independent growth, cell invasion and
128 deregulated mammosphere growth. Phenotypically, the morphology of the MCF10A-CTCF
129 hemizygous cells (CTCF+/-) cells in two dimensional culture is akin to that of CTCF+/+
130 MCF10A cells ([Supplementary Figure 1b](#)). The proliferation rate of the CTCF+/- cells is also
131 comparable to that of the wild-type (WT) cells ([Supplementary Figure 1c](#)) and the cells do not
132 readily form colonies on soft agar. MCF10A cells spontaneously form organized hollow ductal
133 acinar-like structures in three dimensional (3D) culture³⁵ and we next investigated the impact of
134 lowered CTCF levels on growth in this system. Strikingly, CTCF+/- MCF10A acini form
135 significantly larger, less hollow, ($p < 0.0001$) and structurally deformed ([Figure 1b](#),
136 [Supplementary Figure 1d](#)) mammospheres compared to WT counterparts. Subsequently, we
137 performed matrigel transwell invasion assays to investigate the effect of low levels of CTCF on
138 cell invasiveness. The loss of one copy of CTCF allowed MCF10A cell to acquire the capacity to
139 invade through a matrigel matrix. While MCF10A cells harbor little capacity to invade, CTCF+/-
140 cells readily invaded through matrigel at a rate 10-fold higher than WT cells ($p = 0.0066$ and $p <$
141 0.0001 for CTCF+/- #1 and #2) ([Figure 1c](#)). Together, these results point towards a potentially

142 important role for the loss of heterozygosity of CTCF in cancer progression, as it promotes
143 disorganized 3D growth and invasiveness, two strongly linked oncogenic abilities critical for
144 tumors to progress from a benign state to later stages of cancer.

145 **CDH1 loss does not phenocopy CTCF hemizyosity**

146

147 CTCF is located at the chromosomal locus 16q22.1, which is commonly deleted in cancers³⁶.
148 This loss commonly encompasses multiple genes, most notably the tumor suppressor gene
149 CDH1, whose mutation may play an important role in lobular carcinoma of the breast³⁷. To
150 compare CDH1 loss with that of CTCF, we generated CDH1 knockout (KO) MCF10A cells
151 using CRISPR-Cas9 (Supplementary Figure 1e). In stark contrast to CTCF+/- cells, the
152 invasiveness of the CDH1 KO was not increased compared to WT MCF10A (Supplementary
153 Figure 1e). These data validate the relevance and specificity of our CTCF+/- model and
154 strengthens the concept that loss of CTCF is a major oncogenic target of 16q22.1 deletion.

155 To support a direct link between the loss of CTCF and the acquirement of invasiveness, we
156 carried out lentiviral-mediated addback of HA-CTCF within our MCF10A models. As expected,
157 the restoration of CTCF levels was able to reduce the invasiveness of CTCF+/- cells to a level
158 comparable to WT cells (Figure 1d/e). To further explore a link between reduced pools of CTCF
159 and cell invasion, we utilized conditionally reprogrammed cell lines derived from TNBC PDXs
160 with confirmed CTCF LOH (Supplementary Figure 1f) and low CTCF expression (Figure 1f).
161 Following lentiviral addback of HA-CTCF in the PDX cell lines (Figure 1g), the resultant
162 increased expression of CTCF led to a reversal of the invasive phenotype in all three patient-
163 derived cell lines tested (Figure 1h). Together, these results highlight a direct association
164 between CTCF loss of heterozygosity and cancer progression of multiple normal and cancerous
165 breast epithelial models.

166 **Reprogramming of transcriptional networks leads to activation of oncogenic** 167 **signaling in CTCF hemizygous cells**

168

169 To gain insight into the mechanism whereby CTCF+/- cells acquire the oncogenic phenotypes
170 described above, we carried out an RNA-Seq to compare global gene expression profiles of WT
171 MCF10A and CTCF+/- MCF10A cells. Using DESEQ2³⁸, we detected 2976 and 2893 genes that

172 were significantly transcriptionally altered in the CTCF+/- #1 and #2, respectively, compared to
173 WT (Basemean > 100, abs (log2FC) >1, adjusted p-value < 0.05). The transcriptional changes
174 were highly reproducible as 2765 genes were commonly altered in both CTCF+/- clones (
175 approximately 93% of #1 and 96% of #2). Of these genes, a slight majority of 1503 genes were
176 upregulated (54%), compared 1261 genes that were downregulated (46%) (Figure 2a). These
177 results demonstrate that a specific subset of genes is consistently sensitive to CTCF depletion in
178 breast epithelial cells.

179 GSEA analysis of the RNA-Seq data revealed that multiple gene sets related to the
180 phosphoinositide 3-kinase (PI3K) and epithelial to mesenchyme transition (EMT) pathways were
181 strongly upregulated in the CTCF+/- cells (Figure 2b/c). The PI3K pathway is a classical
182 oncogenic pathway aberrantly activated in diverse cancers that drives both invasiveness and
183 altered mammosphere morphology³⁵. Among the top upregulated genes in our CTCF+/- clones
184 were ERBB3 and FGFR1, well-characterized receptor tyrosine kinases and oncogenes that
185 activate the phosphorylation cascade of the PI3K pathway^{39,40}. An important downstream effect
186 of PI3K is the phosphorylation of 4E-BP1. Phosphorylated 4E-BP1 cannot bind the translation
187 initiation factor eIF4E, thereby facilitating the interaction of eIF4E with the translational
188 initiation machinery. Thus, PI3K signaling feeds into the EMT pathway, partially through
189 translational upregulation of classical oncogenes. SNAI1, which is a target of eIF4E-mediated
190 translational upregulation, itself promotes invasion⁴¹. Strikingly, SNAI1 was among the top hits
191 within the EMT pathway based on our RNA-seq data (Figure 2c). The marked upregulation of
192 these top hits was validated using qPCR (Figure 2d). Downregulated genes were enriched for
193 those involved in cellular adhesion (Supplementary Figure 2a/b) pathways, consistent with our
194 invasive phenotype and upregulation of SNAI1⁴². Considering that CTCF+/- cells do not
195 undergo obvious changes in morphology, it is likely that these cells undergo a partial EMT that
196 is reversible upon re-expression of CTCF.

197 **Activation of the PI3K pathway in CTCF hemizygous cells**

198

199 Now guided by our RNA-seq data, we next aimed to examine whether the PI3K pathway was
200 hyper-activated in the MCF10A CTCF+/- cells and if so, determine whether this signaling
201 contributes to the invasive phenotype of these cells. First, we screen for increased activation of

202 key downstream effectors of PI3K signaling, including phosphorylation of 4EBP1 (serine 65)
203 and S6K1 (threonine 389)⁴³. Under conditions of serum starvation, where phosphorylation of
204 S6K1 and 4EBP1 were weakly detected in the WT cells, a strong signal of both S6K1 and
205 4EBP1 phosphorylation, direct targets of the mTORC1 complex, was detected in the CTCF+/-
206 cells (Figure 3a). As the upregulation of the PI3K pathway can alter the morphology of
207 mammospheres³⁵, we tested whether the elevated activation of PI3K might be observed under 3D
208 culture conditions. Similarly, we detected a pronounced phosphorylation of S6, the direct target
209 of S6K1, in both CTCF+/- mammosphere populations, while it was absent in the WT acini
210 (Figure 3b). The outer region of mammospheres is expected to be the primary proliferative zone
211 due to accessibility to nutrients and oxygen. We developed a custom script to visually isolate and
212 quantify the fluorescence intensity of this proliferative region for individual mammospheres. We
213 detected that phosphorylation of S6 was 3.1 and 2.7 times higher in CTCF+/- #1 and #2 than in
214 the WT acini ($p < 0.0001$). Thus, under conditions of both 2D and 3D culture, reduced pools of
215 CTCF leads to transcriptional reprogramming that activates the PI3K pathway.

216 Based on the aberrant activation of PI3K signaling in CTCF+/- cells, we surmised that their
217 invasivity may be vulnerable to inhibitors of this pathway. To test this hypothesis, we targeted
218 mTORC1/2 because these kinase complexes assimilate the signals from diverse branches of the
219 PI3K signaling cascade⁴³. We carried out matrigel transwell invasion assays following a 24h
220 starvation and 48h mTORC1/2 inhibition, using the second generation mTor inhibitor Torin1⁴⁴.
221 Since the WT MCF10A are mostly non-invasive, we compared the changes in invasiveness of
222 our CTCF hemizygous cells to the well-characterized triple negative breast cancer cell line,
223 MDA-MB-231. Following Torin1 treatment at 10nM and 25nM, the MDA-MB-231 invasiveness
224 was largely unaffected. However, the CTCF KOs were markedly sensitive to these low Torin1
225 concentrations, as they were significantly blocked in their ability to invade (Figure 3c). These
226 data indicate that the PI3K pathway plays a central role in driving the invasion of normal
227 epithelial cells with reduced CTCF levels, while late stage TNBC lines, such as MDA-MB-231
228 may utilize multiple, or alternative, pathways to achieve this phenotype.

229 We also validated that low concentrations of Torin1 treatment efficiently resolved the activation
230 of mTORC1. Concentrations as low as 5nM strongly abrogate the phosphorylation of 4EBP1 in
231 the CTCF+/- cells under starved condition (Figure 3d). As the PI3K pathway has also been

232 shown to control the protein expression of SNAI1 through translational upregulation⁴¹, we
233 investigated the impact of Torin1 treatment on SNAI1 expression. We detected a marked, dose-
234 dependent drop in SNAI1 protein levels following 24h of Torin1 exposure (Figure 3d). Since
235 SNAI1 overexpression promotes invasiveness in multiple models^{45,46} and it is strongly
236 overexpressed at the mRNA and protein level in our CTCF+/- MCF10A (Figure 3e), we decided
237 to investigate whether it is an important downstream target of PI3K and playing a role in the
238 invasiveness of the CTCF+/- cells. To do so, we used lentiviral-mediated shRNA knockdown of
239 SNAI1 and surveyed the changes to cell invasion. The downregulation of SNAI1 led to both a
240 significant reduction of SNAI1 protein levels and of CTCF+/- invasiveness (Figure 3f).
241 Interestingly, the association between high SNAI1 expression and low CTCF is also observed
242 clinically. We probed TCGA Breast Cancer data and found a significant correlation between low
243 CTCF and high SNAI1 expression in patients' breast tumors (Supplementary Figure 3a). Overall,
244 these results highlight the importance of the upregulation of the PI3K pathway, and its
245 downstream effector, SNAI1, for the oncogenicity of the CTCF+/- cells. These indicate that the
246 invasion of tumors harboring CTCF copy number loss, coupled with elevated SNAI1, may be
247 susceptible to therapeutic intervention with inhibitors of mTORC1/2.

248 **Altered CTCF binding to DNA surrounding oncogenes in hemizygous cells**

249

250 While CTCF+/- cells are viable, it was reasonable to expect that the reduced nuclear pool of
251 CTCF would likely compromise the number of occupied CTCF sites on chromatin. As a first
252 step towards understanding the mechanisms through which CTCF+/- cells alters its
253 transcriptome, and upregulates oncogenic transcriptional networks, we carried out a ChIP-Seq to
254 map CTCF binding across the genome (Figure 4a). We identified that a majority of CTCF
255 binding sites, 38 775 out of the 44 802 peaks called by MACS2⁴⁷, were left unchanged between
256 the WT and CTCF+/- MCF10A. Considering that CTCF levels were reduced by ~50-60% in
257 these cells, it is clear that the nuclear pool of CTCF is in excess of that required for genomic
258 regulation, consistent with previous reports showing a significant fraction of CTCF is unbound
259 within interphase cell populations⁴⁸. This excess is likely a safeguard against genomic instability
260 and protection of the transcriptome that might stem from fluctuating CTCF levels. However, as
261 expected, a subset of 5313 sites displayed reduced or lost CTCF binding in the CTCF+/-
262 MCF10As compared to the WT (FDR < 0.01, LogFC < -1). Surprisingly, a small cluster of 714

263 sites displayed a gain of CTCF binding in the CTCF \pm cells compared to WT (FDR < 0.01,
264 LogFC > 1) (Figure 4b).

265 We continued this analysis by investigating the differences in binding strength and distribution
266 between the sites lost and those remaining constant in CTCF \pm cells. The average read density
267 was lower for DNA binding sites within the lost cluster compared to the constant cluster in the
268 WT MCF10As (Figure 4c). The genomic distribution of lost sites was also unique compared to
269 constant sites, with lost sites being enriched at promoter regions. Twenty-nine percent of CTCF
270 lost sites compared to 19% of constant sites, were found on promoters, such as the promoter of
271 SNAI1 (Supplementary Figure 4a). Intergenic CTCF were more consistent, with 37% of the lost
272 sites, compared to 41% of constant sites, being found at distal intergenic regions, including a
273 ERBB3 downstream site (Figure 4d, Supplementary Figure 4b). These results suggests that a
274 subset of weakly binding CTCF sites, mildly enriched at promoter regions, are more sensitive to
275 CTCF depletion.

276 Consistent with our RNA-Seq, KEGG pathway analysis of the CTCF binding sites in the lost
277 cluster were strongly enriched surrounding genes involved in the PI3K-Akt signaling pathway
278 (ranked 2nd by relative gene count), as well as multiple pathways related to cell mobility, such as
279 focal adhesion (ranked 3rd) and ECM-receptor activation (ranked 4th) (Figure 4e). In summary,
280 in normal breast epithelial cells, lost sites of CTCF are likely to occur at weaker elements, which
281 are markedly enriched at genes involved in the PI3K pathway and cell invasion, such as SNAI1.

282 **CTCF Lost Sites are frequently proximal to deregulated genes**

283

284 CTCF may impact gene transcription through binding proximal or distal to transcription start
285 sites (TSS) through multiple mechanisms. To gain insight into the mechanisms whereby altered
286 CTCF binding might impact transcriptional events in the MCF10A hemizygous cells, we
287 mapped lost sites to determine their proximity to TSS. About half of all CTCF lost sites (2408
288 out of the 5313) were found with proximity (\pm 3kb) to significantly altered genes (Basemean >
289 100, adjusted p-value < 0.05). Interestingly, a significant fraction of which are found around
290 strongly upregulated (Log2FC > 1, 530 sites) and strongly downregulated (Log2FC < -1, 716
291 sites) genes (Figure 4f). Therefore, the lost sites of CTCF are in general located with relative
292 proximity to loci transcriptionally altered. The intensity of the loss of CTCF significantly, but

293 weakly, correlated with both upregulation ($r = -0.1$, $p = 0.0056$) and downregulation ($r = 0.2$, $p <$
294 0.0001) of gene expression. Similarly, both upregulated and downregulated genes displayed a
295 slightly lower average CTCF ChIP-Seq read density on their TSS (Figure 4g). Overall, proximal
296 loss of CTCF may be associated with both upregulation and downregulation of gene expression,
297 but only mildly correlates with the intensity of altered gene expression. This hints that in many
298 cases, the changes in transcription observed in CTCF hemizygous cells are likely driven through
299 indirect mechanisms including changes in chromatin conformation or epigenetic reprogramming
300 but not due to a loss of CTCF interactions with core transcription machinery^{49,50}.

301 **CTCF loss potentiates epigenetic reprogramming at transcriptionally altered genes** 302

303 Destabilization of CTCF binding has been linked to numerous epigenetic changes and it is likely
304 that the insulator function of CTCF contributes to epigenetic patterning of the genome⁵¹⁻⁵³. Thus,
305 we next investigated whether the changes to gene expression and CTCF binding were associated
306 with changes to chromatin marks. First, we screened for multiple activating and silencing histone
307 marks on representative altered genes using ChIP-qPCR. Although we detected significant
308 changes in H3K4me3 and H2K27ac associated with altered transcription (Supplementary Figure
309 5a/b), we did not detect strong changes with the repressive marks H3K27me3 and H3K9me3
310 (Supplementary Figure 5c/d), which is consistent with a previous study where changes to CTCF
311 binding across multiple genomes were not strongly linked to differences in H3K27me3⁵².
312 Therefore, we proceeded to map H3K4me3 and H3K27ac genome-wide using ChIP-seq to
313 compare WT MCF10A with CTCF \pm cells (Figure 5a). A majority of H3K4me3 (~82%) and
314 H3K27ac (~87%) peaks displayed similar read counts and were conserved between CTCF \pm and
315 the WT MCF10A. However, both H3K4me3 and H3K27ac showed significant alterations upon
316 loss of CTCF. H3K4me3 and H3K27ac gained enrichment at 2929 and 5188 loci respectively.
317 Further, H3K4me3 was reduced at 1932 and H3K27ac at 2060 sites ($\text{abs}(\text{LogFC}) \geq 1$, $\text{FDR} \leq$
318 0.05) (Figure 5b). Overall, CTCF loss potentiated a gain of marks associated with gene
319 activation. Then, we aimed to assess whether these changes to histone marks were likely driving
320 the changes in gene expression. We observed a pronounced, and statistically significant gain of
321 H3K27ac at upregulated genes ($r = 0.64$, $p < 0.0001$) (Figure 5c) including oncogenes such as
322 ERBB3 and SNAIL (Figure 5d), compared to a more modest correlation of H3K4me3 with
323 upregulated genes ($r = 0.45$, $p < 0.0001$) (Figure 5c).

324 CTCF loss has been indirectly linked to deregulated DNA methylation^{54,55} and it is possible that
325 altered DNA methylation contributes to transcriptomic changes observed in CTCF+/- cells. We
326 carried out bisulfite conversion and investigated the association between genome wide changes
327 in DNA methylation, using Illumina EPIC methyl array with our RNA-Seq data. Contrarily to
328 the strong correlation detected between changes in activating marks and gene expression, the
329 changes in DNA methylation pattern observed in the CTCF+/- did not correlate with changes in
330 gene expression ($r = -0.04$, $p < 0.0001$) (Figure 5e). These results indicate that under conditions of
331 sub-physiological CTCF levels, changes in gene expression most specifically linked to a global
332 reprogramming of H3K27ac.

333 To test for a role of gained H3K27ac in the promotion of cell invasion, we treated CTCF+/- cells
334 with the Histone Acetyl-Transferase (HAT) inhibitor A485, that targets CBP⁵⁶. Under serum
335 starved conditions, A485 treatment efficiently resolves the hyperactivation of the PI3K/mTor
336 pathway in these cells, indicated by a dose-dependent reduction of 4EBP1 phosphorylation
337 (Figure 5f). Similarly, CBP inhibition blocked SNAI1 expression, linking the gain of H3K27ac
338 to its upregulation (Figure 5f). As inhibition of both the PI3K pathway and SNAI1 expression
339 reduced the invasiveness of the CTCF+/- cells, we tested their invasivity after exposure to A485
340 treatment and further, compared the effects with those observed in MDA-MB-231 cells.
341 Similarly to mTor inhibition and SNAI1 knockdown, A485 treatment significantly reduced the
342 invasiveness of the CTCF+/- cells (Figure 5g) further supporting the hypothesis that the
343 increased deposition of H3K27ac plays a key role in the oncogenic phenotypes caused by the
344 loss of CTCF. Interestingly, MDA-MB-231 cells were noticeably sensitive to this treatment
345 (Figure 5g) as well. These results highlight a general dependency on increased histone
346 acetylation during the invasion process of aggressive epithelial cancer cells and supports an
347 essential role of epigenetic reprogramming during cancer progression.

348 **Reduced CTCF Levels lead to loss of insulation of subTAD structures**

349

350 Following our ChIP-Seq experiments, we posited that the loss of CTCF binding and the relative
351 increase in open chromatin at activated genes may stem from a loss of insulation. Therefore, we
352 aimed to investigate changes in 3D chromatin architecture using Hi-C. Here, we generated 600
353 million reads per condition with biological replicates of the WT, CTCF+/- #1 and #2 merged for

354 high resolution analysis. This sequencing depth allowed us to reach a complete genomic
355 coverage at 5kb resolution, consistent with previous high-resolution Hi-C data⁵⁷⁻⁵⁹. Genome-
356 wide, the megabase scale chromatin interactions in our WT cells ([Supplementary Figure 6a](#))
357 were strikingly consistent with previously published data⁶⁰. At the megabase scale, we also did
358 not detect notable changes in chromosome organization between WT and CTCF+/- cells
359 ([Supplementary Figure 6a](#)).

360 Next, we queried whether more local changes chromatin architecture may underlie the RNA
361 profiles resulting from CTCF CNL. First, to detect and quantify TAD changes at the sub-
362 megabase scale, we used a hierarchical TAD caller, hiTAD⁶¹, to call TAD boundaries and
363 domain boundaries within TADs (termed subTADs) at a 10kb resolution. We then compared the
364 colocalization of called boundaries (+/- 10kb) between the WT and two CTCF+/- clones. Of the
365 11,580 TAD boundaries called, 10% were lost in both CTCF+/- lines compared to the WT.
366 These changes were more pronounced when looking at subTAD boundaries, where 17% of the
367 total number were lost ([Figure 6a](#)). The loss of these boundaries might potentiate de novo
368 contacts between DNA elements due to loss of insulation. Indeed, CTCF+/- cells gained 810 new
369 TAD boundaries (7% gain) and 606 subTAD boundaries (11% gain), which are enriched next to
370 lost boundaries ([Supplementary Figure 6b](#)), indicating a re-organization of sub-genomic regions.
371 Altered boundaries frequently colocalized with altered sites of CTCF binding (+/- 10kb), with
372 lost boundaries showing a marked enrichment for lost CTCF elements, while gained boundaries
373 are generally CTCF-null ([Figure 6b](#)). These de novo TAD/subTAD interactions, demarcated by
374 gained boundaries, are likely generated from a loss of insulator activity that limit long range
375 DNA contacts, so it is logical that these regions would be devoid of CTCF. This novel
376 mechanism is supported by a recent study by Vos et al. demonstrating that CTCF-independent
377 enhancer looping is potentiated by the loss of proximal CTCF binding⁶².

378 To validate that the loss of CTCF binding leads to local loss of DNA insulation, we imaged the
379 average local interaction centered around lost sites of CTCF (+/-200kb, [Figure 6c](#)). In agreement
380 with our hypothesis, we detected a marked reduction of boundary strength, represented by a
381 decreased interaction intensity at CTCF sites delimiting two domains. DNA insulation was also
382 clearly compromised as represented by an increased interaction intensity between the domains
383 spanning the lost sites of CTCF ([Figure 6c](#)). Subsequently, we asked whether loss of insulation

384 was equally concurrent with loss of TADs and subTADs. To answer this question, we subdivided
385 the lost CTCF sites into lost sites colocalizing with a TAD boundary or alternatively, grouped
386 lost sites found within TADs (subTADs). For each subdivision of the lost sites, we plotted local
387 interactions centered around the lost sites of CTCF (+/-200kb) and measured the average
388 insulation score of these regions using FAN-C⁶³. We detected a slight loss of boundary strength
389 and insulation at lost sites of CTCF colocalizing with TAD boundaries. However, lost sites of
390 CTCF within TADs resulted in a nearly complete loss of boundary strength and insulation,
391 allowing inter-domain DNA interactions (Figure 6d). These observations validate, in a
392 quantifiable manner, the prominent loss of insulation at subTADs under conditions of CTCF
393 hemizyosity.

394 **Changes in subTAD organization drives epigenetic reprogramming and changes in** 395 **gene expression** 396

397 We continued our Hi-C analysis to investigate whether the changes in subTAD interactions are
398 connected to the changes in gene expression. First, we measured the average gene expression
399 changes at altered subTAD and TAD boundaries. Genes colocalizing with the gained subTAD
400 boundaries were the most significantly upregulated ($p < 0.0001$, Figure 6e) compared to all
401 genes. Interestingly, altered TAD boundaries were not significantly associated to transcriptional
402 changes (Figure 6e). Importantly, the genes found at gained boundaries were enriched for genes
403 involved in the PI3K pathway (Supplementary Figure 6c). Overall, these analysis reveal that
404 reorganization of subTADs may lead to aberrant activation of the PI3K pathway.

405 We find that altered subTAD interactions and changes to activating marks are both associated
406 with changes in gene expression, so we investigated whether colocalization of H3K27ac or
407 H3K4me3 was observed at domain boundaries. We detected a strong enrichment of gained sites
408 of H3K27ac and H3K4me3 with gained subTAD and vice-versa with lost subTADs (Figure 6f,
409 Supplementary Figure 6d). However, altered TAD boundaries were not enriched for changes in
410 either mark (Figure 6f, Supplementary Figure 6d) consistent with the lack of transcriptional
411 changes in these regions. These results are validated by comparing the average changes in
412 insulation at gained sites of H3K27ac colocalizing with TAD boundaries or within TADs (Figure
413 6g). Gain of H3K27ac at TAD boundaries was not predictive of changes of TAD insulation,

414 while gain of H3K27ac within TADs led to a marked gain of insulation (Figure 6g), confirming
415 the formation of de novo subTAD boundaries at these sites. Again using pileup plots, we looked
416 at the average density of interactions between regions of gained H3K27ac and all sites of either
417 H3K27ac or H3K4me3 (Figure 6h). Considering all combinations, we detected a marked gain of
418 interaction at loci where H3K27ac was gained in the CTCF[±] cells (Figure 6h). These results
419 indicate that the reconfiguration of subTADs allows for de novo interactions at regulatory
420 regions enriched for gains of H3K27ac that drive the expression of oncogenic programs.

421 An excellent example of this mechanism may be observed at the SNAI1 locus. At the megabase
422 scale conformational changes are not obvious (Supplementary Figure 7a). Using HIFI⁶⁴ to
423 facilitate Hi-C resolution at a sub-5K scale at the SNAI1 locus, we detected a discrete, novel
424 interaction between the SNAI1 gene and a downstream enhancer in CTCF[±] cells (Figure 7a).
425 This interaction is positioned adjacent to the lost CTCF binding site within the SNAI1 promoter,
426 and is embedded with a region of gained H3K27ac (Figure 7b). The downstream enhancer,
427 connecting with the promoter, is likewise enriched for H3K27ac in the CTCF[±] cells (Figure
428 7b).

429 To validate that the loss of CTCF at SNAI1 may drive its overexpression, we directed a dCAS9
430 construct to the CTCF site at SNAI1 promoter (sgSNAI1), in MCF10A WT cells, where CTCF
431 binding is compromised in CTCF[±] cells. Disruption of CTCF at this site with exogenous
432 dCAS9 would be expected to facilitate an increase of SNAI1 expression if our model is correct.
433 As a control we used a sgRNA targeting a CTCF-unbound region at the SNAI1 locus (sgCTL).
434 Compared to WT cells infected with sgCTL, cells infected with sgSNAI1 displayed a significant
435 upregulation of SNAI1 mRNA levels (2.1 fold increase, $p = 0.006$) (Figure 7c). As a further
436 control, directing dCAS9 to this CTCF binding site in hemizygous cells, where CTCF binding is
437 already compromised, did not result in an upregulation of SNAI1 (Figure 7c). These data
438 validate that disruption of CTCF may play a key role in driving the upregulation of oncogenes,
439 including SNAI1.

440 In summary, the loss boundaries at the subTAD level compromises insulation from de novo
441 contacts. These de novo contacts, and the associated enrichment for H3K27ac at these regions, in
442 turn, play a major role in driving the oncogenic networks observed in CTCF hemizygous cells.

443 **Discussion**

444

445 In this study, we demonstrated for the first time that a global loss of insulation of subTAD
446 domains, caused by reduced pools of CTCF may promote breast cancer progression. This loss of
447 insulation leads to shifts in subTAD boundaries which were strongly associated with increased
448 deposition of the activating histone marks, H3K27ac and H3K4me3, and transcriptional
449 upregulation of oncogenes of the PI3K pathway and SNAI1. As well as advancing our
450 knowledge on the physiological importance of proper topological organization of chromatin
451 within TADs, this information highlights a therapeutic sensitivity to mTor inhibitors for tumors
452 with low CTCF expression or altered subTAD insulation.

453 The PI3K pathway is one of the most aberrantly activated across cancers. Activation of PI3K by
454 oncogenic signals such as oxidative stress⁶⁵ leads to translational upregulation of proliferative
455 and survival factors such as SNAI1⁴¹. While the transcriptional regulation of oncogenic signaling
456 pathways is not fully understood, we now show that reorganization of subTADs may play an
457 important role in the control of oncogenic signaling. PI3K signaling is influenced by the many
458 factors including growth factors, oxidative stress and nutrient availability. As such, it would be
459 logical to expect that transcriptional control of PI3K signaling would be likewise impacted by
460 regulatory mechanisms that respond quickly to similar environmental influences. Dynamic
461 remodeling of subTAD organization represents one such mechanism.

462 Most TAD boundaries are often bound by multiple redundant CTCF binding sites and colocalize
463 with tRNA or housekeeping genes that are constitutively active^{66,67}. Unlike TADS, which are
464 generally consistent across tissue types, and in our study, very tolerant to reduced pools of
465 CTCF, subTADs reorganize themselves to allow for dynamic and precise transcriptional control
466 of genomic loci⁶⁸. Consistent with this concept, long range chromatin interactions are re-
467 organized during serum starvation, dependent upon interactions between CTCF and binding
468 partners⁶⁹ and a gain of CTCF-mediated interactions at the subTAD level have been correlated
469 with gene expression⁷⁰. Indeed, CTCF may play a central role in dynamic reorganization of
470 subTADs, with promoter-bound CTCF serving as a docking site for differential enhancer
471 contacts⁶⁸. However, based on a model of subTAD reorganization that is partially dependent on a
472 redistribution of CTCF binding, this binding should be influenced by changes to the cellular

473 environment. CTCF sites are known to be variable across tissue types, dependent on DNA
474 methylation that compromises the binding of a subset of CTCF sites⁵⁵ and emerging data
475 suggests that CTCF binding profiles are not static but react to the cellular milieu. Under
476 conditions of cellular stress in breast cancer cells, CTCF is targeted for phosphorylation,
477 reducing its affinity for DNA at select genes involving those in PI3K and mTor signaling⁷¹. In
478 the context of cancer, our study suggests that the loss of CTCF association with weak binding
479 elements frequently leads to a loss of insulation resulting in de novo enhancer promoter
480 interactions that activate a PI3K transcriptional program.

481 Our study indicates that CTCF LOH is one means through which CTCF binding is compromised
482 but clearly additional mechanisms are relevant. While infrequent, CTCF mutations are common
483 across cancers and are enriched in endocrine therapy resistant tumors⁷². CTCF binding sites
484 themselves are frequently mutated across cancers including those proximal to SNAIL in gastric
485 cancer⁷³. Based on our data that compromised CTCF plays a role in cancer progression, it is
486 tempting to speculate that signaling events that correspond with tumor mobility and invasion
487 may compromise CTCF binding further accelerating metastasis. LATS signaling influences
488 CTCF binding to chromatin during stress⁷¹ and it would be worthwhile to investigate whether
489 other processes that facilitate metastasis such as hypoxia or cytoskeletal remodeling also
490 influence the genomic distribution of CTCF.

491 Our data also indicate that the development of a therapeutic approach aimed at restoration of
492 insulation at specific genomic sites could prove useful to prevent metastasis. While such an
493 approach may be challenging, we suggest that tumors defined by reorganization of long range
494 contacts may be susceptible to mTor inhibitors and HATi inhibitors.

495 The histone modification H3K27ac is commonly used to demarcate contact between
496 transcriptionally active regions of the genome. Indeed, multiple HiChIP experiments have
497 validated a link between enhancers and promoters marked by H3K27ac and the presence of
498 chromatin loops^{74,75}. Further, combined with CTCF ChIP-Seq data, H3K27ac ChIP-Seq data has
499 been used to infer the three-dimensional organization of DNA, without performing Hi-C
500 experiments⁷⁶, highlighting the strong relationship between H3K27ac, CTCF binding, and
501 chromatin conformation. However, it remains unclear how chromatin contacts and histone
502 modifications influence each other. Our data provides insight into the relationship between

503 altered subTAD distribution and epigenetic changes. Since sites of compromised CTCF binding
504 are generally not proximal to sites of gained H3K27ac or H3K4me3, it is unlikely that the loss of
505 CTCF drives the gain of activating marks through direct mechanisms. For example, it is unlikely
506 that compromised CTCF binding would lead to the loss of recruitment of antagonistic epigenetic
507 writers, such as EZH2⁷⁷ at sites many kilobases away where H3K27ac is subsequently
508 accumulated. Therefore, we can infer that the reshuffling of subTADs is driving the
509 redistribution of H3K27ac and H3K4me3 more so than altered CTCF itself, by allowing novel
510 chromatin contacts between genomic regions. In turn, these contacts may promote the
511 recruitment of activating chromatin writers^{13,14}, leading to reprogrammed epigenetic landscape
512 and transcriptional changes.

513 We also highlighted the importance of histone acetylation in the oncogenic process of invasion.
514 Therefore, epigenetic reprogramming events, potentially driven by subTAD reorganization,
515 could underlie aggressive tumor behavior. Such gain of interactions might also occur
516 independently of CTCF status. These observations hint at the potential therapeutic effectiveness
517 of histone acetyl transferase inhibitor to block the epigenetic shifts necessary for tumor
518 progression.

519 Overall, our study highlights the interplay between long range chromatin contacts and epigenetic
520 remodeling and underscores the importance of defining epigenetic reprogramming in cancer as a
521 means to uncover novel therapeutic avenues.

522

523 **Methods**

524 **Cell Culture**

525 MCF10A cell line and mutants were maintained in DMEM/F12 50/50 (Wisent, #319-085-CL)
526 supplemented with EGF (100µg/ml, Wisent, #511-110-UM), Insulin (10mg/ml, Wisent, #H511-
527 016-U6), Hydrocortisone (1mg/ml, Sigma, #H0888-1G), Horse Serum (2%, Wisent, #065150) and
528 Cholera toxin (1mg/ml, Sigma, #C8052-2MG) in an incubator at 37C and 5% CO₂.

529 MDA-MB-231 and HEK293T cell lines were maintained in DMEM (Wisent, #319-005-CL)
530 supplemented with 10% FBS (Gibco, #12483-020) in an incubator at 37C and 5% CO₂.

531 Conditionally Reprogrammed Cells of Patients Derived Xenograft of Triple Negative Breast
532 Cancer tumors, termed PDXs, were established following as in Liu et al. 2017⁷⁸ and Sirois et al.
533 2019⁷⁹ and given to us by Dr. Park's and Dr. Basik's Laboratories from McGill University. Low
534 CTCF protein expression in these cell lines was confirmed by Western Blot. Loss of
535 Heterozygosity was confirmed using the Chromosome Analysis Suite from Thermo Fisher. They
536 were maintained in DMEM (Wisent, #319-005-CL), with 25% Ham's F12 (Wisent, #312-250-
537 CL), 8% FBS (Gibco, #12483-020), L-Glutamine (1.5mM, Wisent, #609-065-EL), EGF (50µg/ml,
538 Wisent, #511-110-UM), Insulin (5mg/ml, Wisent, #H511-016-U6), Hydrocortisone (0.8mg/ml,
539 Sigma, #H0888-1G), Cholera toxin (84µg/ml, Sigma, #C8052-2MG), RhoK Inhibitor (0.01mM,
540 Y-27623, StemCell Technologies, #72304) in an incubator at 37C and 5% CO₂.

541 **CRISPR-Cas9 Knock-Out**

542 MCF10A CTCF+/- cells lines were previously used in Hilmi et al 2017⁸⁰. CRISPR/Cas9 KO of
543 CDH1 in MCF10A was performed as in Hilmi et al 2017⁸⁰. In brief, 1 million MCF10A cells are
544 seeded in a 6cm dish and transfected the next day using Lipofectamine 3000 (Invitrogen, #
545 L3000001), 6µg of pCas-Guide-ef1a-GFP (OriGene, #GE100018, guide sequence :
546 ACGTCGTTACGAGTCACTTCAGG). Two days later, GFP-positive cells were isolated by
547 fluorescence-activated cell sorting into 96-well plates (one cell per well). To screen for CDH1 KO
548 clones, we isolated genomic DNA of each clone and amplified proximal sequences surrounding
549 the Cas9 targets by polymerase chain reaction. Positive clones were first identified using the
550 SURVEYOR Assay Kit (IDT, #706020) followed by Sanger Sequencing at Genome Quebec
551 facilities.

552

553 **Western Blot**

554 Western blots were carried out as previously described⁸⁰. For Western Blot conducted on PDX
555 tumors, tissue was harvested as in Savage et al 2020⁸¹. For all other Western Blots, cells were
556 harvested by scrapping. Then, cells are lysed in whole-cell lysis buffer [20 mM tris (pH 7.5), 420
557 mM NaCl, 2 mM MgCl₂, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.5% Triton X-100,
558 supplemented with fresh 1 mM dithiothreitol, phenylmethylsulfonyl fluoride, protease inhibitor
559 cocktail (Roche) and phosphatase inhibitors, bis-glycerol phosphate, and NaF] for 15min, then
560 spined at 13000rpm at 4°C for 15min to pellet cellular debris. Then, the protein concentration of
561 the supernatant is assessed using a Bradford assay (Fisher, #1856209). 40μg of proteins are loaded
562 on an 8 to 12% acrylamide gel and electrophoresed at 120V for 1h. Then, proteins on the gel are
563 transferred on nitrocellulose membrane (Pall, #66485) at 4°C, 34V overnight for 8% gel and 100V
564 for 1 hour for 12% gel. The membrane is then blocked with 5% milk in TBST [20 mM Tris base,
565 137 mM NaCl, and 0.1% Tween 20] for 3h at 4°C. The membrane is then incubated with primary
566 antibodies (see [Supplementary Table 1](#)) overnight at 4°C. Membranes are rinsed and washed for
567 10 minutes twice with TBST prior to secondary antibody incubation with goat anti-rabbit
568 (SeraCare, #5220-0458) or anti-mouse (SeraCare, #5450-0011) dilute 1/10000 or 1/20000 in 5%
569 milk in TBST. Membranes are washed again 3 times for 10 minutes in TBST, then revealed using
570 ECL (Bio-Rad, #170-5061). Band intensity quantification and normalization on background is
571 performed using ImageJ software.

572

573 **Growth curves**

574 15000 cells were plated in one well of a 12 wells petri dish (Fisher Scientific, #3513). Cells were
575 fixed at day 1, day 3, and day 5 with 4% formaldehyde and stored at 4°C. Once all the conditions
576 are fixed, the cells are stained with 1ml of a crystal violet solution [1% crystal violet, 10% EtOH]
577 and dried. The stained cells are then diluted in 10% acetic acid. The growth ration of cells is
578 calculated by reading and the comparing the DO at 595nm, the intensity of the violet dye, using
579 PerkinElmer Multimode Plate Reader.

580

581 **Lentiviral Infection for CTCF Addback and SNAI1 Knockdown**

582 CTCF addback (Genecopoeia, #EX-Z8806-Lv120), SNAI1 shRNA knockdown (Sigma,
583 #NM_005985, target sequence: GCAAATACTGCAACAAGGAAT) and GFP control vector
584 (Genecopoeia, #EX-EGFP-Lv120) were done using lentiviral vectors packages in HEK293T cells,
585 as described previously⁸⁰. HEK293T cells were transfected with 7µg of the required lentiviral
586 vector combined with 5µg of packaging vector MD2G and 2µg of envelope vector Pax2 using
587 polyethylenimine (1 mg/ml). 24h after transfection, media was changed for the culture media used
588 for maintenance. Viruses were collected at 48h and 72h after transfection and passed through a
589 0.45-µm filter. For the infection, MCF10A and PDXs cells were infected in six-well dishes and
590 incubated with 1ml of viral supernatant along with 1ml of their respective culture media and
591 8µg/ml of hexadimethrine bromide (Polybrene) for MCF10A and 60µg/ml for PDXs. 24h after
592 infection, culture media is changed and puromycin selection starts.

593 MCF10A cells were selected with 1 µg/ml of puromycin for the first two days following infection,
594 followed by 0.25µg/ml of puromycin for 2 to 3 more days. Culture media is changed for puromycin
595 free media 24h before the starvation period preceding the invasion assay. For the PDX cells,
596 1µg/ml of puromycin is used on the first day of selection, 0.5µg/ml of puromycin is used for the
597 second day, culture media is changed for puromycin-free media on the third day and the
598 preliminary starvation starts on the fourth day.

599

600 **Transwell Invasion Assay and Quantification**

601 70% confluent cells were starved for 24h in non supplemented DMEM/F12, for the MCF10A cells,
602 or DMEM for MDA-MB-231 and PDXs. 50 000 MCF10A cells or 100 000 MDA-MB-231 cells
603 or 200 000 PDX cells are seeded into an insert (Falcon, #353182) coated with 25µg/ml matrigel
604 (Corning, #354230) for the MCF10A and MDA-MB-231 or 20µg/ml matrigel for the PDXs,
605 diluted in in a 0.01M Tris and 0.7M NaCl solution. The cells are maintained in non supplemented
606 media in the insert. The inserts are then placed in companion plate chambers (Falcon, #353503)
607 containing supplemented media used for cell culture overnight for the MCF10A and MDA-MB-
608 231 or for 24h for the PDX.

609 For Torin1 (Tocris, #4247) or A485 (Tocris, #6387) treated conditions, cells were treated with the
610 indicated concentration of Torin1 (0nM, 10nM, 25nM, 50nM) or A485 (0uM, 2uM and 5uM),

611 dilute in DMSO, for 24h before starvation, during starvation and during invasion, in both the insert
612 media and the companion plate media.

613 Then, the inserts are washed in PBS, fixed in 5% glutaraldehyde for 10 minutes, stained with a
614 crystal violet solution [1% crystal violet, 10% EtOH] for 30 minutes, rinsed in water and dried.
615 For each biological replicates, 2 to 3 inserts are plated; for each insert, 5 pictures are taken, at 10X
616 resolution. The total number of invading cells on each picture are then counted using ImageJ
617 software and the average number of invasive cells per 5 pictures per inserts are averaged within
618 each sample and compared between samples. Statistical test between samples is performed using
619 Student t-test.

620 **Lentiviral Infection for dCAS9**

621 20k MCF10A cells are plated in each well of a 6 well-plate. To increase the rate of infection,
622 cells are infected sequentially over a period of 4 days: 24h and 72h after seeding, cells are
623 infected with dCAS9+blasticidin resistance lentiviral construct (Addgene, #85417), with a ratio
624 of media to dCAS9 viral media of 1:1 and 30ug/ul of polybrene; 48h and 96h after seeding, cells
625 are infected with guideRNAbackbone+puromycin resistance lentiviral construct (Addgene,
626 #52963, in which gRNA sequence (CACCGGAGGACAGAGAGACAAGTGT) generated with
627 CHOPCHOP⁸² were cloned into by Norclone), with a ratio of media to gRNA viral media of 1:1
628 and 30ug/ul of polybrene. Culture media is changed every 24h during infection period and
629 afterward, until cells are harvested. 2ug/ul blasticidin selection starts after 48h and is reduced to
630 1ug/ul after 72h, until cells are harvested. 1ug/ul puromycin selection starts after 72h and is
631 reduced to 0.5ug/ul after 96h, until cells are harvested. Cells are harvested for RNA extraction 3
632 days after the last infection (7 days after seeding).

633 **Mammosphere assay and Quantification**

634 5000 cells were seeded on a 50µl matrigel cushion (10-12mg/ml, Corning, #354230) and
635 maintained in supplemented DMEM/F12 containing 4% matrigel for 8 days. The media is
636 carefully replaced every 3 days. Average mammosphere size was measured from brightfield
637 microscopy images on ImageJ software. Statistical test between the average mammosphere size
638 of each sample was performed using Student t-test.

639 **Mammosphere Immunofluorescence and Quantification**

640 p-S6 immunofluorescence was performed using p-S6 S240/244 antibody from Cell Signaling
641 (Rabbit, #2215S) and Goat Anti-Rabbit IgG with Alexa 488 fluorophore (Invitrogen, #A32731).
642 DAPI was used for DNA fluorescence of the whole mammosphere used for normalization of p-S6
643 fluorescence quantification and mammosphere filling quantification.

644 Mammosphere filling was quantified from Z-stacks of DAPI stained mammosphere images using
645 ImageJ software. The ratio between the area of the hollow cavity and the total area of the
646 mammosphere was measured on each Z-stacks of each mammosphere of each sample and the Z-
647 stack with the highest ratio was selected and quantified for each mammosphere of each sample.
648 Statistical test between the filling ratio of samples was performed using Student t-test.

649 The quantification of p-S6 fluorescence was performed using a custom script in ImageJ developed
650 by Dr. Luke McCaffrey's group. In brief, mammospheres were detected by thresholding the image
651 (Mean method) to create a whole-organoid mask. This mask was duplicated then iteratively eroded
652 (13 times) to create an inner mask that excluded the outer layer of cells. A mask for the outer layer
653 of cells was generated using an XOR gate applied to the whole organoid and inner mask. The mean
654 pixel intensity (8-bit) was measured under the mask, for each whole organoid, outer, and inner
655 regions. The mean pixel intensity of each region was then compared between the samples.
656 Statistical test between the p-S6 outer fluorescence of each sample was performed using Student
657 t-test.

658

659 **RNA-Seq**

660 **RNA-Seq Sample preparation and sequencing**

661 Total RNA was extracted according to Sigma RNA Extraction Kit (Sigma, #RTN350-1KT)
662 protocol. RNA quantity and quality was measured using Nanodrop. RNA was sent to Genome
663 Quebec for polyA RNA library preparation using NEBNext Ultra II Directional RNA Library Prep
664 Kit for Illumina and sequencing of 50M 100bp Paired-End reads per replicate on Illumina
665 NovaSeq 6000 platform.

666 **RNA-Seq Data Processing and Analysis**

667 The overall quality of reads and sequencing was assessed before and after trimming using the
668 FastQC package (Babraham Bioinformatics). Prior to mapping, reads were trimmed with

669 Trimmomatics⁸³ using the following condition: ILLUMINACLIP:\$Adapters:2:30:10:8:true,
670 HEADCROP:4,SLIDINGWINDOW:4:30,LEADING:3,TRAILING:3,MINLEN:30. Alignment
671 on hg19 human genome was performed with STAR 2.5.4b⁸⁴ default parameters, and converted
672 into bam format using Samtools 1.9⁸⁵. Differential expression analysis was generated using
673 FeatureCounts count matrix⁸⁶ followed by DESEQ2 analysis³⁸, using default parameters and
674 prefiltering, for comparison across samples.

675 **RNA-Seq Volcano Plot**

676 Volcano plot representation of the RNA-Seq results was generated using the DESEQ2 calculated
677 Log2FC and -log(adjusted p-value) of the respective MCF10A CTCF+/- compared to MCF10A
678 WT for every gene with a basemean > 100. Genes with p-value < 0.05 were represented in black.
679 Genes with Log2FC > 0 were represented in orange. Genes with Log2FC < 0 were represented in
680 purple.

681 **RNA-Seq GSEA Pathway Analysis**

682 Pathway analysis was performed using GSEA tools default setting on the read count matrix of all
683 significantly altered genes (basemean > 100, p.value <0.05)⁸⁷. All gene sets shown were significant
684 for both p-value (< 0.001) and FDR (<0.25). Pathway names were shortened as follows, with the
685 full name of each pathway being:

686 **PI3K Signaling** : GO_PHOSPHATIDYLINOSITOL_3_KINASE_SIGNALING,

687 **EMT** : GO_EPITHELIAL_TO_MESENCHYMAL_TRANSITION,

688 **EMT Regulation** : GO_POSITIVE_REGULATION_OF_EPITHELIAL_CELL_MIGRATION,

689 **PI3K Regulation** : GO_POSITIVE_REGULATION_OF_PHOSPHATIDYLINOSITOL_3_KINASE_SIGNALING

690 **Cell-Cell Adhesion** :

691 GOBP_HETEROPHILIC_CELL_CELL_ADHESION_VIA_PLASMA_MEMBRANE_CELL_ADHESION_MOLECULES

692 **RNA-Seq Heatmaps**

693 Heatmaps were generated using the Log2FC of the 36 genes with the highest absolute Log2FC in
694 the PI3K Regulation and EMT Regulation genesets.

695

696 **RT-qPCR**

697 Total RNA was extracted according to Sigma RNA Extraction Kit (Sigma, #RTN350-1KT)
698 protocol. RNA quantity and quality was measured using Nanodrop. 500ng of RNA are used as
699 template for Reverse-Transcriptase PCR, following the manufacturer protocol (All-In-One RT
700 MasterMix, ABM, #G490). The cDNA is diluted 1:10 and 2µl is used for qPCR amplification,

701 following the manufacturer protocol (GoTaq qPCR MasterMix 2X, Promega, #A600A). Relative
702 levels of cDNA are compared between the samples using the $2^{-\Delta\Delta CT}$ formula normalized on the
703 average level of 3 housekeeping genes (GapDH, RPL4 and RPLPO). Statistical test between
704 normalized gene expression of each gene for each sample is performed using Student t-test. The
705 sets of primers used for RT-qPCR are listed in [Supplementary Table 2](#).

706

707 **TCGA Data Analysis**

708 The RNA-Seq data of Breast Cancer Patients from the 2018 TCGA Dataset (generated by the
709 TCGA Research Network: <https://www.cancer.gov/tcga>) was used. This dataset was separated into
710 the high CTCF group, being the top 20% of patients in term of highest CTCF expression, and the
711 low CTCF group, being the top 20% of patients in term of low CTCF expression. The two groups
712 were compared for SNAI1 RNA expression using a Student's T Test and a Spearman correlation
713 test.

714 **ChIP-Seq**

715 **ChIP-Seq Sample preparation and sequencing**

716 70-80% confluent cells were fixed 10 minutes in 4% formaldehyde and stored at -80C. The pellets
717 were subsequently resuspended in 1ml of ChIP-buffer [0.25% NP-40, 0.25% Triton X-100, 0.25%
718 Sodium Deoxycholate, 0.005% SDS, 50nM Tris (pH8), 100mM NaCl, 5mM EDTA, 1X PMSF,
719 2mM NaF, 1X P8340 Cocktail Inhibitor (Roche)] and sonicated with a probe sonicator (Fisher
720 Scientific Sonic Dismembrator Model 500) using the following cycles: 5 cycles at 20% power, 5
721 cycles at 25% power, and 5 cycles at 30% power. Each cycle lasts 10 seconds, and the samples are
722 kept on ice between each cycle to avoid overheating. Next, the samples are spun at high speed in
723 a microcentrifuge for 30 minutes. Then, lysates are collected and protein concentration measured
724 using the Bradford assay, as described above. Based on protein concentrations, samples are diluted
725 to 2mg/ml proteins in ChIP-buffer and 50ul/ml of Protein G Plus-Agarose Suspension Beads
726 (Calbiochem, IP04-1.5ML) are added for 3h to preclear. 2% of the sample is collected as input and
727 kept at -20 °C until DNA purification. Immunoprecipitation is carried out at 4°C overnight with
728 1ml of sample, 60ul of beads and primary antibody (see [Supplementary Table 1](#)). The beads are
729 then washed once with Wash1, Wash2, Wash3 [0.10% SDS, 1% Triton X-100, 2mM EDTA,
730 20mM Tris (pH 8), 150/200/500mM NaCl for Wash 1,2,3 respectively], Wash LiCl [0.25M LiCl,

731 1% NP-40, 1% Sodium Deoxycholate, 1mM EDTA, 10mM Tris (pH8)] and twice with TE buffer
732 [10mM Tris (pH8), 1mM EDTA]. Then, beads are resuspended in elution buffer [1% SDS, 0.1M
733 NaHCO₃]. The samples are decrosslinked overnight at 65 °C. 20µg of Proteinase K (Sigma, #
734 39450-01-6) is added for 1h at 42 °C. Then, DNA is purified using BioBasic DNA collection
735 column (BioBasic, #SD5005). DNA concentration was assessed via Picogreen assay (Invitrogen,
736 #P7589). Minimally, 15ng of each ChIP sample, with 2 biological replicates and one input per cell
737 lines (beside the ChIP-Seq on CTCF in MCF10A WT where 4 biological replicates were
738 generated), was sent to Genome Quebec (for CTCF and H3K4me3) or The Centre for Applied
739 Genomics (TCAG) at SickKids Hospital (for H3K27ac) for library preparation using NEB Ultra
740 II DNA kit (no shearing) and next-generation sequencing on Illumina Platform (Single-End 100bp
741 sequencing, ~50M reads per sample).

742 **ChIP-Seq Data Processing and Analysis**

743 Quality control of reads and sequencing was assessed before and after trimming by FastQC
744 (Babraham Bioinformatics). Reads were trimmed with Trimmomatics⁸³ using the following
745 parameters: ILLUMINACLIP:\$Adapters:2:30:10, LEADING:30, TRAILING:30,
746 SLIDINGWINDOW:4:30, MINLEN:30. Alignment on hg19 human genome was performed using
747 BWA⁸⁵ default conditions. Sam files generated by BWA were converted to bam format using
748 Samtools⁸⁵. Peak calling was performed with MACS2⁴⁷ default condition and normalization on
749 the respective Input dataset of each cell lines. Bigwig files used for visualization were generated
750 from the fragment pileup bedGraph using the BedGraphToBigwig function.

751 **ChIP-Seq Differential Binding Analysis**

752 Differentially binding region were quantified using DiffBind 3.0⁸⁸. Bam and narrowPeak files for
753 each samples and bam files of the corresponding input were used. Default normalization and
754 analysis was performed for H3K4me3 and H3K27ac. CTCF normalization and analysis was
755 performed with the following parameters: normalize = DBA_NORM_DEFAULT, library =
756 DBA_LIBSIZE_PEAKREADS, background = F, bREtrieve = F. Threshold of significance were
757 set at FDR <= 0.01 and abs(LogFC) >= 1 in all conditions. Consensus differential peaksets
758 between replicates and conditions were used for further downstream analysis and converted to
759 Grange format using GenomicRanges R packages⁸⁹. The number of sites in each peak set was used
760 for quantification and the generation of pie chart.

761 **ChIP-Seq Genomic Distribution and Sites Annotation**

762 Genomic distribution and annotation were performed using clusterProfiler package⁹⁰ and
763 ChIPSeeker package⁹¹ on the differential binding sites identified by DiffBind 3.0, using the
764 TxDb.Hsapiens.UCSC.hg19.knownGene as reference for gene location. TSS regions were defined
765 with a +/- 3000bp overlap during peak annotation. The regions are annotated as : 5'UTR, Promoter
766 (≤ 1 kb), Promoter (1-2kb), Promoter (2-3kb) are referred to as "Promoter (+/- 3kb)"; 1st Exon and
767 Other Exon are referred to as "Exons"; 1st Intron and Other Intron are referred to as "Introns";
768 3'UTR and Downstream are referred to as "Downstream" and Distal Intergenic is referred to as is.

769 **ChIP-Seq KEGG Pathway Analysis**

770 Pathway analysis was performed using the annotation files from above and using the
771 compareCluster function from the previously mentioned clusterProfiler package and the following
772 parameters: geneCluster = genes, fun = "enrichKEGG", pvalueCutoff = 0.05, pAdjustMethod =
773 "BH". $-\log(p.value)$ of enrichment significance was used for bar chart representation, where
774 pathways were ranked according to geneRatio.

775 **ChIP-Seq and RNA-Seq Dot Plots and Correlation**

776 Dot plots were made by combining the RNA-Seq Log2FC between WT and CTCF+/- #2 and the
777 logFC from DiffBind of any called peak annotated on that gene (+/- 3kb) by clusterProfiler.
778 Spearman correlation on the dot plot were performed using the ChIP-Seq logFC and RNA-Seq
779 log2FC of every peak colocalization with a gene. Genes associated with a peak with a DiffBind
780 or DESEQ2 adjusted p-value > 0.05 or DiffBind LogFC or DESEQ2 Log2FC < 1 are represented
781 in grey, while those with a LogFC/ Log2FC ≥ 1 or ≤ -1 are represented in orange and purple,
782 respectively.

783 **ChIP-Seq Heatmaps, Profile Plot, Tracks**

784 Heatmaps, profile plot and tracks were generate using deepTools and samtools^{85,92}. Heatmaps and
785 Profile plot were generated using 3kb regions centered around the differential peakset identified
786 by DiffBind and bigwig from MACS2. Both the computeMatrix and plotHeatmaps were runned
787 with default parameter; yMax, zMax and colors were adjusted in each condition to better represent
788 the results. Tracks were generated as profile plot of the single genomic regions of interest with a
789 gene annotation track from IGV⁹³ under each figure to represent the relative location of the gene
790 of interest.

791 **ChIP-Seq Colocalization Analysis**

792 Analysis of colocalization, +/- 3kb, was performed using a genomic overlap algorithm between
793 the position of differentially binding peak sets identified by DiffBind. Observed/Expected ratios
794 shown for colocalization of ChIP-Seq peaks were calculated using the Chi-Square formula in
795 Microsoft Excel.

796

797 **ChIP-qPCR**

798 The Chromatin Immunoprecipitation was done following the ChIP-Seq protocol, however using
799 only 1mg/ml of chromatin and 30ul of beads with the antibodies listed in [Supplementary Table 1](#).
800 Final ChIP-product is diluted in 60ul of DNase-free water. qPCR was performed with the ChIP
801 product following the manufacturer protocol (GoTaq qPCR MasterMix 2X, Promega, #A600A).
802 $2^{-\Delta\Delta CT}$ formula was used for quantification, normalized on a 2% chromatin input of each sample
803 and compared between sites and conditions. Primers used for ChIP DNA amplification are
804 documented in [Supplementary Table 3](#).

805

806 **EPIC Array Investigation**

807 **Bisulfite conversion and Methyl Array**

808 Bisulfite conversion was performed using EZ DNA Methylation Kit (Zymo Research, #D5001).
809 500ng per sample of Bisulfite-converted DNA was sent at Princess Margaret Genomics Centre for
810 quality control and detection of methylated bases using Illumina Human Methylation EPIC Array.

811 **Methyl Array Data Analysis**

812 The “.idat” files outputted from the Illumina EPIC Array experiment were analyzed using Minfi
813 package⁹⁴ for the comparison of individual red/green CpG probe intensity and genomic annotation,
814 using Illumina Methylation EPIC reference: ilm10b4.hg19. Quality control of methylation pattern
815 was performed using the Shinymethyl R package⁹⁵.

816 **Methyl Array and RNA-Seq Dot Plot**

817 Dot plot of methylation profile and RNA-Seq was generated using the Methylation LogFC
818 between MCF10A CTCF+/- #2 and MCF10A WT for all CpG colocalizing with a gene (+/- 2kb)
819 with a DESEQ2 called basemean > 100. Spearman correlation was calculated using all points
820 under these criteria.

821

822 **Hi-C**

823 **Hi-C Sample preparation and sequencing**

824 Hi-C data was generated from 1M WT, CTCF+/- #1 or CTCF+/- #2 MCF10A cells per
825 replicates, with 2 replicates per conditions, using the Arima-HiC kit, according to the
826 manufacturer's protocols (Arima Genomics). Library preparation was performed using KAPA
827 Hyper Prep Kit (#07962312001) following Arima protocol for library preparation. Libraries
828 were sent at TCAG at SickKids Hospital for next-generation sequencing using Illumina
829 Novaseq S1 flowcell (Paired-End 50bp Sequencing, ~300M reads per replicate)

830 **Hi-C Data Processing and Analysis**

831 Quality control of reads and sequencing was assessed by FastQC (Babraham Bioinformatics). Raw
832 sequencing read were mapped, filtered, and binned using the runHiC pipeline⁶¹. Contact matrix
833 were binned at 5kb and 10kb resolution and stored in “.cool” format.

834 **Hierarchical TAD Calling**

835 Hierarchical TAD calling was performed using the hiTAD function of the TADLib package⁶¹,
836 using the 10kb resolution contact matrix and default settings.

837 **Domain Boundaries Colocalization Analysis**

838 Colocalization of TAD boundaries and ChIP-Seq peak was determine as described earlier, using a
839 simple genomic overlap algorithm between the called TAD boundaries and the differentially
840 bound peak list generated with DiffBind, with an accepted overlap of +/- 10kb (+/- 1 contact matrix
841 bin). Observed/Expected calculation were performed as described earlier. The same algorithm
842 and overlap were used across samples to determine altered boundaries. TAD and subTAD
843 boundaries present only in the WT, but in none of both the CTCF+/- clones were defined and
844 quantified as lost boundaries. TAD and subTAD boundaries present in both CTCF+/- clones, but
845 absent in the WT were defined and quantified as gained boundaries. Boundaries found in the WT
846 and any of the CTCF+/- clone were defined as constant boundaries.

847 Constant or gained boundaries were defined as adjacent to lost boundaries is the next or second to
848 next boundaries, in any direction, is lost. Significant enrichment of the gained boundaries next to
849 lost boundaries, compared to constant boundaries, was performed using the Chi-Square formula
850 in Microsoft Excel.

851 **RNA-Seq Changes at Domain Boundaries**

852 Constant and altered boundaries' genomic location (10kb each) were annotated as previously
853 described in the ChIP-Seq Analysis section. The average RNA-Seq log2FC of the CTCF+/- #2
854 against the WT of genes colocalizing with each type of boundaries was calculated and represented
855 by a bar chart. Statistical test of the difference between the average RNA Log2FC at each boundary
856 was performed using a Student's T Test. Pathway analysis of altered boundaries was performed as
857 in the ChIP-Seq section.

858 **Hi-C 2D Heatmaps**

859 Genome wide and Chromosome 20 heatmaps were generated using Juicer⁹⁶ representation of
860 observed interacting reads value. 5kb resolution heatmaps were generated from h5 converted cool
861 files using the HiCExplorer packages⁹⁷ default settings, using the hicPlotMatrix function for
862 interaction matrix and hicCompareMatrices for comparative interaction matrix and
863 hicConvertFormat for the cool to h5 conversion. High resolution sub5kb HiC imaging was
864 performed with the HIFI pipeline⁶⁴, using the Markovian Recombination Method on defined
865 subsection of the genome (around SNAI1, chr20:48,550,000-48,750,000) and default setting for
866 this method.

867 **Hi-C Pile-Up Plots**

868 Pile-Up plots were generated using the cooltools package (Open Chromosome Collective),
869 centered at the differential peak list from DiffBind (for CTCF and histone marks) or from the
870 TADlib TAD caller (for TAD/subTAD boundaries), normalized on random background
871 interaction and using default settings. Local interactions were map at +/- 200kb around the defined
872 regions. Average interactions were mapped at +/-50kb around the defined region.

873 **Insulation Score and Profile Plots**

874 Insulation Score was calculated at 30kb resolution and outputted as a bigwig file using the FAN-
875 C insulation command default settings⁶³ and the respective 10kb .cool matrix from MCF10A WT
876 and MCF10A CTCF+/- #2. The bigwig file was used to generate the profile plot using deepTools,
877 previously explained in the ChIP-Seq section, at +/- 200kb around lost sites of CTCF or gained
878 sites of H3K27ac colocalizing with TAD boundaries (+/- 10kb) or within TADs.

879 **QUANTIFICATION AND STATISTICAL ANALYSIS**

880

881 Unless stated otherwise, all graphical representations display the mean and SEM of the sample's
882 distribution. Unless stated otherwise, graphics and statistical tests were generated and performed
883 using GraphPad Prism 9.1, GraphPad Software, San Diego, California USA,
884 www.graphpad.com. Graphical abstract was created using Biorender.

885 For each graph:

886 * : $p \leq 0.05$

887 ** : $p \leq 0.01$

888 *** : $p \leq 0.001$

889 **** : $p < 0.0001$

890

891

892 **References:**

893

- 894 1 Chen, C. *et al.* Spatial Genome Re-organization between Fetal and Adult Hematopoietic
895 Stem Cells. *Cell Rep* **29**, 4200-4211 e4207, doi:10.1016/j.celrep.2019.11.065 (2019).
- 896 2 Sivakumar, A., de Las Heras, J. I. & Schirmer, E. C. Spatial Genome Organization: From
897 Development to Disease. *Front Cell Dev Biol* **7**, 18, doi:10.3389/fcell.2019.00018
898 (2019).
- 899 3 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals
900 folding principles of the human genome. *Science* **326**, 289-293,
901 doi:10.1126/science.1181369 (2009).
- 902 4 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of
903 chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 904 5 Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345,
905 doi:10.1126/science.aaz3418 (2019).
- 906 6 Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA
907 by loop extrusion. *Science* **366**, 1345-1349, doi:10.1126/science.aaz4475 (2019).
- 908 7 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324,
909 doi:10.1016/j.cell.2017.09.026 (2017).
- 910 8 Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and
911 Enhancer/Promoter Function. *Cell* **162**, 900-910, doi:10.1016/j.cell.2015.07.038 (2015).
- 912 9 Hsieh, T. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian
913 Chromatin Folding. *Mol Cell* **78**, 539-553 e538, doi:10.1016/j.molcel.2020.03.002
914 (2020).
- 915 10 Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of
916 genomes during lineage commitment. *Cell* **153**, 1281-1295,
917 doi:10.1016/j.cell.2013.04.053 (2013).
- 918 11 Tena, J. J. & Santos-Pereira, J. M. Topologically Associating Domains and Regulatory
919 Landscapes in Development, Evolution and Disease. *Front Cell Dev Biol* **9**, 702787,
920 doi:10.3389/fcell.2021.702787 (2021).
- 921 12 Matthews, B. J. & Waxman, D. J. Computational prediction of CTCF/cohesin-based
922 intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver.
923 *Elife* **7**, doi:10.7554/eLife.34077 (2018).
- 924 13 Kang, H. *et al.* Dynamic regulation of histone modifications and long-range
925 chromosomal interactions during postmitotic transcriptional reactivation. *Genes Dev* **34**,
926 913-930, doi:10.1101/gad.335794.119 (2020).
- 927 14 Greenwald, W. W. *et al.* Subtle changes in chromatin loop contact propensity are
928 associated with differential gene regulation and expression. *Nat Commun* **10**, 1054,
929 doi:10.1038/s41467-019-08940-5 (2019).
- 930 15 Chen, X. *et al.* Key role for CTCF in establishing chromatin structure in human embryos.
931 *Nature* **576**, 306-310, doi:10.1038/s41586-019-1812-0 (2019).
- 932 16 Justice, M., Carico, Z. M., Stefan, H. C. & Downen, J. M. A WIZ/Cohesin/CTCF
933 Complex Anchors DNA Loops to Define Gene Expression and Cell Identity. *Cell Rep* **31**,
934 107503, doi:10.1016/j.celrep.2020.03.067 (2020).

935 17 Luo, H. *et al.* HOTTIP lncRNA Promotes Hematopoietic Stem Cell Self-Renewal
936 Leading to AML-like Disease in Mice. *Cancer Cell* **36**, 645-659 e648,
937 doi:10.1016/j.ccell.2019.10.011 (2019).

938 18 Willemin, A. *et al.* Induction of a chromatin boundary in vivo upon insertion of a TAD
939 border. *PLoS Genet* **17**, e1009691, doi:10.1371/journal.pgen.1009691 (2021).

940 19 Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of
941 enhancer-promoter looping. *Nucleic Acids Res* **47**, 6699-6713, doi:10.1093/nar/gkz462
942 (2019).

943 20 Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat*
944 *Genet* **47**, 818-821, doi:10.1038/ng.3335 (2015).

945 21 Guo, Y. *et al.* CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop
946 anchors identifies repressive chromatin loops. *Genome Biol* **19**, 160, doi:10.1186/s13059-
947 018-1531-0 (2018).

948 22 Guo, Y. A. *et al.* Mutation hotspots at CTCF binding sites coupled to chromosomal
949 instability in gastrointestinal cancers. *Nat Commun* **9**, 1520, doi:10.1038/s41467-018-
950 03828-2 (2018).

951 23 Korkmaz, G. *et al.* A CRISPR-Cas9 screen identifies essential CTCF anchor sites for
952 estrogen receptor-driven breast cancer cell proliferation. *Nucleic Acids Res* **47**, 9557-
953 9572, doi:10.1093/nar/gkz675 (2019).

954 24 Bailey, C. G. *et al.* CTCF Expression is Essential for Somatic Cell Viability and
955 Protection Against Cancer. *Int J Mol Sci* **19**, doi:10.3390/ijms19123832 (2018).

956 25 Moore, J. M. *et al.* Loss of maternal CTCF is associated with peri-implantation lethality
957 of Ctf null embryos. *PLoS One* **7**, e34915, doi:10.1371/journal.pone.0034915 (2012).

958 26 Alharbi, A. B. *et al.* Ctf haploinsufficiency mediates intron retention in a tissue-specific
959 manner. *RNA Biol* **18**, 93-103, doi:10.1080/15476286.2020.1796052 (2021).

960 27 Kemp, C. J. *et al.* CTCF haploinsufficiency destabilizes DNA methylation and
961 predisposes to cancer. *Cell Rep* **7**, 1020-1029, doi:10.1016/j.celrep.2014.04.004 (2014).

962 28 Aitken, S. J. *et al.* CTCF maintains regulatory homeostasis of cancer pathways. *Genome*
963 *Biol* **19**, 106, doi:10.1186/s13059-018-1484-3 (2018).

964 29 Damaschke, N. A. *et al.* CTCF loss mediates unique DNA hypermethylation landscapes
965 in human cancers. *Clin Epigenetics* **12**, 80, doi:10.1186/s13148-020-00869-7 (2020).

966 30 Driouch, K. *et al.* Loss of heterozygosity on chromosome arm 16q in breast cancer
967 metastases. *Genes Chromosomes Cancer* **19**, 185-191, doi:10.1002/(sici)1098-
968 2264(199707)19:3<185::aid-gcc8>3.0.co;2-u (1997).

969 31 Schmidt-Kittler, O. *et al.* From latent disseminated cells to overt metastasis: genetic
970 analysis of systemic breast cancer progression. *Proc Natl Acad Sci U S A* **100**, 7737-
971 7742, doi:10.1073/pnas.1331931100 (2003).

972 32 Thomassen, M., Tan, Q. & Kruse, T. A. Gene expression meta-analysis identifies
973 chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast*
974 *Cancer Res Treat* **113**, 239-249, doi:10.1007/s10549-008-9927-2 (2009).

975 33 Cattoglio, C. *et al.* Determining cellular CTCF and cohesin abundances to constrain 3D
976 genome models. *Elife* **8**, doi:10.7554/eLife.40164 (2019).

977 34 Tian, T., Li, X. & Zhang, J. mTOR Signaling in Cancer and mTOR Inhibitors in Solid
978 Tumor Targeting Therapy. *Int J Mol Sci* **20**, doi:10.3390/ijms20030755 (2019).

979 35 Lourenco, C. *et al.* Modelling the MYC-driven normal-to-tumour switch in breast cancer.
980 *Dis Model Mech* **12**, doi:10.1242/dmm.038083 (2019).

981 36 Filippova, G. N. *et al.* A widely expressed transcription factor with multiple DNA
982 sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of
983 the smallest regions of overlap for common deletions in breast and prostate cancers.
984 *Genes Chromosomes Cancer* **22**, 26-36 (1998).

985 37 Petridis, C. *et al.* Germline CDH1 mutations in bilateral lobular carcinoma in situ. *Br J*
986 *Cancer* **110**, 1053-1057, doi:10.1038/bjc.2013.792 (2014).

987 38 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
988 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8
989 (2014).

990 39 Kiavue, N. *et al.* ERBB3 mutations in cancer: biological aspects, prevalence and
991 therapeutics. *Oncogene* **39**, 487-502, doi:10.1038/s41388-019-1001-5 (2020).

992 40 Servetto, A., Formisano, L. & Arteaga, C. L. FGFR signaling and endocrine resistance in
993 breast cancer: Challenges for the clinical development of FGFR inhibitors. *Biochim*
994 *Biophys Acta Rev Cancer* **1876**, 188595, doi:10.1016/j.bbcan.2021.188595 (2021).

995 41 Robichaud, N. *et al.* Phosphorylation of eIF4E promotes EMT and metastasis via
996 translational control of SNAIL and MMP-3. *Oncogene* **34**, 2032-2042,
997 doi:10.1038/onc.2014.146 (2015).

998 42 Haraguchi, M. *et al.* Snail regulates cell-matrix adhesion by regulation of the expression
999 of integrins and basement membrane proteins. *J Biol Chem* **283**, 23514-23523,
1000 doi:10.1074/jbc.M801125200 (2008).

1001 43 Dibble, C. C. & Cantley, L. C. Regulation of mTORC1 by PI3K signaling. *Trends Cell*
1002 *Biol* **25**, 545-555, doi:10.1016/j.tcb.2015.06.002 (2015).

1003 44 Thoreen, C. C. *et al.* An ATP-competitive mammalian target of rapamycin inhibitor
1004 reveals rapamycin-resistant functions of mTORC1. *J Biol Chem* **284**, 8023-8032,
1005 doi:10.1074/jbc.M900301200 (2009).

1006 45 Jung, H. Y. *et al.* Apical-basal polarity inhibits epithelial-mesenchymal transition and
1007 tumour metastasis by PAR-complex-mediated SNAIL degradation. *Nat Cell Biol* **21**, 359-
1008 371, doi:10.1038/s41556-019-0291-8 (2019).

1009 46 Li, C. *et al.* Propofol facilitates migration and invasion of oral squamous cell carcinoma
1010 cells by upregulating SNAIL expression. *Life Sci* **241**, 117143,
1011 doi:10.1016/j.lfs.2019.117143 (2020).

1012 47 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
1013 doi:10.1186/gb-2008-9-9-r137 (2008).

1014 48 Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin
1015 regulate chromatin loop stability with distinct dynamics. *Elife* **6**, doi:10.7554/eLife.25776
1016 (2017).

1017 49 Liu, Z., Scannell, D. R., Eisen, M. B. & Tjian, R. Control of embryonic stem cell lineage
1018 commitment by core promoter factor, TAF3. *Cell* **146**, 720-731,
1019 doi:10.1016/j.cell.2011.08.005 (2011).

1020 50 Pena-Hernandez, R. *et al.* Genome-wide targeting of the epigenetic regulatory protein
1021 CTCF to gene promoters by the transcription factor TFII-I. *Proc Natl Acad Sci U S A*
1022 **112**, E677-686, doi:10.1073/pnas.1416674112 (2015).

1023 51 Witcher, M. & Emerson, B. M. Epigenetic silencing of the p16(INK4a) tumor suppressor
1024 is associated with loss of CTCF binding and a chromatin boundary. *Mol Cell* **34**, 271-
1025 284, doi:10.1016/j.molcel.2009.04.001 (2009).

1026 52 Fang, C. *et al.* Cancer-specific CTCF binding facilitates oncogenic transcriptional
1027 dysregulation. *Genome Biol* **21**, 247, doi:10.1186/s13059-020-02152-7 (2020).

1028 53 Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox
1029 clusters during differentiation. *Science* **347**, 1017-1021, doi:10.1126/science.1262088
1030 (2015).

1031 54 Soochit, W. *et al.* CTCF chromatin residence time controls three-dimensional genome
1032 organization, gene expression and DNA methylation in pluripotent cells. *Nat Cell Biol*
1033 **23**, 881-893, doi:10.1038/s41556-021-00722-w (2021).

1034 55 Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation.
1035 *Genome Res* **22**, 1680-1688, doi:10.1101/gr.136101.111 (2012).

1036 56 Lasko, L. M. *et al.* Discovery of a selective catalytic p300/CBP inhibitor that targets
1037 lineage-specific tumours. *Nature* **550**, 128-132, doi:10.1038/nature24028 (2017).

1038 57 Ren, B. *et al.* High-resolution Hi-C maps highlight multiscale 3D epigenome
1039 reprogramming during pancreatic cancer metastasis. *J Hematol Oncol* **14**, 120,
1040 doi:10.1186/s13045-021-01131-0 (2021).

1041 58 Bauer, M. *et al.* Chromosome compartments on the inactive X guide TAD formation
1042 independently of transcription during X-reactivation. *Nat Commun* **12**, 3499,
1043 doi:10.1038/s41467-021-23610-1 (2021).

1044 59 Wang, J. *et al.* Epigenomic landscape and 3D genome structure in pediatric high-grade
1045 glioma. *Sci Adv* **7**, doi:10.1126/sciadv.abg4126 (2021).

1046 60 Barutcu, A. R. *et al.* Chromatin interaction analysis reveals changes in small chromosome
1047 and telomere clustering between epithelial and breast cancer cells. *Genome Biol* **16**, 214,
1048 doi:10.1186/s13059-015-0768-0 (2015).

1049 61 Wang, X. T., Cui, W. & Peng, C. HiTAD: detecting the structural and functional
1050 hierarchies of topologically associating domains from chromatin interactions. *Nucleic
1051 Acids Res* **45**, e163, doi:10.1093/nar/gkx735 (2017).

1052 62 Vos, E. S. M. *et al.* Interplay between CTCF boundaries and a super enhancer controls
1053 cohesin extrusion trajectories and gene expression. *Mol Cell* **81**, 3082-3095 e3086,
1054 doi:10.1016/j.molcel.2021.06.008 (2021).

1055 63 Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the
1056 analysis and visualisation of chromosome conformation capture data. *Genome Biol* **21**,
1057 303, doi:10.1186/s13059-020-02215-9 (2020).

1058 64 Cameron, C. J., Dostie, J. & Blanchette, M. HIFI: estimating DNA-DNA interaction
1059 frequency from Hi-C data at restriction-fragment resolution. *Genome Biol* **21**, 11,
1060 doi:10.1186/s13059-019-1913-y (2020).

1061 65 Zhang, T. *et al.* Targeting the ROS/PI3K/AKT/HIF-1alpha/HK2 axis of breast cancer
1062 cells: Combined administration of Polydatin and 2-Deoxy-d-glucose. *J Cell Mol Med* **23**,
1063 3711-3723, doi:10.1111/jcmm.14276 (2019).

1064 66 Iwasaki, Y., Ikemura, T., Kurokawa, K. & Okada, N. Implication of a new function of
1065 human tDNAs in chromatin organization. *Sci Rep* **10**, 17440, doi:10.1038/s41598-020-
1066 74499-7 (2020).

1067 67 Cubenas-Potts, C. *et al.* Different enhancer classes in *Drosophila* bind distinct
1068 architectural proteins and mediate unique chromatin interactions and 3D architecture.
1069 *Nucleic Acids Res* **45**, 1714-1730, doi:10.1093/nar/gkw1114 (2017).

1070 68 Oh, S. *et al.* Enhancer release and retargeting activates disease-susceptibility genes.
1071 *Nature* **595**, 735-740, doi:10.1038/s41586-021-03577-1 (2021).

1072 69 Ferrari, R. *et al.* TFIIC Binding to Alu Elements Controls Gene Expression via
1073 Chromatin Looping and Histone Acetylation. *Mol Cell* **77**, 475-487 e411,
1074 doi:10.1016/j.molcel.2019.10.020 (2020).

1075 70 Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125-129,
1076 doi:10.1038/s41586-021-03639-4 (2021).

1077 71 Luo, H. *et al.* LATS kinase-mediated CTCF phosphorylation and selective loss of
1078 genomic binding. *Sci Adv* **6**, eaaw4651, doi:10.1126/sciadv.aaw4651 (2020).

1079 72 Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast
1080 Cancers. *Cancer Cell* **34**, 427-438 e426, doi:10.1016/j.ccell.2018.08.008 (2018).

1081 73 Umer, H. M. *et al.* A Significant Regulatory Mutation Burden at a High-Affinity Position
1082 of the CTCF Motif in Gastrointestinal Cancers. *Hum Mutat* **37**, 904-913,
1083 doi:10.1002/humu.23014 (2016).

1084 74 Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target
1085 genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612,
1086 doi:10.1038/ng.3963 (2017).

1087 75 Wang, C. *et al.* Primary effusion lymphoma enhancer connectome links super-enhancers
1088 to dependency factors. *Nat Commun* **11**, 6318, doi:10.1038/s41467-020-20136-w (2020).

1089 76 Buckle, A., Brackley, C. A., Boyle, S., Marenduzzo, D. & Gilbert, N. Polymer
1090 Simulations of Heteromorphic Chromatin Predict the 3D Folding of Complex Genomic
1091 Loci. *Mol Cell* **72**, 786-797 e711, doi:10.1016/j.molcel.2018.09.016 (2018).

1092 77 Wei, L. *et al.* Knockdown of CTCF reduces the binding of EZH2 and affects the
1093 methylation of the SOCS3 promoter in hepatocellular carcinoma. *Int J Biochem Cell Biol*
1094 **120**, 105685, doi:10.1016/j.biocel.2020.105685 (2020).

1095 78 Liu, X. *et al.* Conditional reprogramming and long-term expansion of normal and tumor
1096 cells from human biospecimens. *Nat Protoc* **12**, 439-451, doi:10.1038/nprot.2016.174
1097 (2017).

1098 79 Sirois, I. *et al.* A Unique Morphological Phenotype in Chemoresistant Triple-Negative
1099 Breast Cancer Reveals Metabolic Reprogramming and PLIN4 Expression as a Molecular
1100 Vulnerability. *Mol Cancer Res* **17**, 2492-2507, doi:10.1158/1541-7786.MCR-19-0264
1101 (2019).

1102 80 Hilmi, K. *et al.* CTCF facilitates DNA double-strand break repair by enhancing
1103 homologous recombination repair. *Sci Adv* **3**, e1601898, doi:10.1126/sciadv.1601898
1104 (2017).

1105 81 Savage, P. *et al.* Chemogenomic profiling of breast cancer patient-derived xenografts
1106 reveals targetable vulnerabilities for difficult-to-treat tumors. *Commun Biol* **3**, 310,
1107 doi:10.1038/s42003-020-1042-x (2020).

1108 82 Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome
1109 editing. *Nucleic Acids Res* **47**, W171-W174, doi:10.1093/nar/gkz365 (2019).

1110 83 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
1111 sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

1112 84 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
1113 doi:10.1093/bioinformatics/bts635 (2013).

1114 85 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
1115 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

1116 86 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
1117 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
1118 doi:10.1093/bioinformatics/btt656 (2014).

1119 87 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
1120 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-
1121 15550, doi:10.1073/pnas.0506580102 (2005).

1122 88 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical
1123 outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730 (2012).

1124 89 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS*
1125 *Comput Biol* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

1126 90 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing
1127 biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118
1128 (2012).

1129 91 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak
1130 annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383,
1131 doi:10.1093/bioinformatics/btv145 (2015).

1132 92 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible
1133 platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191,
1134 doi:10.1093/nar/gku365 (2014).

1135 93 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26,
1136 doi:10.1038/nbt.1754 (2011).

1137 94 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the
1138 analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369,
1139 doi:10.1093/bioinformatics/btu049 (2014).

1140 95 Fortin, J. P., Fertig, E. & Hansen, K. shinyMethyl: interactive quality control of Illumina
1141 450k DNA methylation arrays in R. *F1000Res* **3**, 175,
1142 doi:10.12688/f1000research.4680.2 (2014).

1143 96 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution
1144 Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).

1145 97 Ramirez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome
1146 organization in flies. *Nat Commun* **9**, 189, doi:10.1038/s41467-017-02525-w (2018).

1147

1148

1149 **Acknowledgements**

1150 This research was supported by the Canadian Institutes of Health Research (CIHR) (159759 and
1151 159663 to M.W. An Oncopole EMC2 grant supported M.W, M.B. and M.P. B.L. is supported by
1152 a Lady Davis Institute TD Bank Studentship Award and M.W. is supported by a Fonds de
1153 Recherche du Québec Santé (FRQ-S), Senior Chercheur Boursier award. The authors would like
1154 to thank Dr. Ivan Topisirovic for insightful comments and editing. We further thank Kathlein
1155 Klein, of the Lady Davis bioinformatics core facility, for her valuable input.

1156

1157

1158

1159

1160

1161 **Author Contributions:**

1162

1163 Conceptualization, B.L., M.J. and M.W. ; Methodology, B.L., M.J. and M.W. ; Software, B.L.,
1164 M.J., S.H., R.C., L.M. and C.L.; Validation, B.L., M.J., T.Z. and N.W.; Formal Analysis, B.L.
1165 and M.J.; Investigation, B.L., M.J., T.J., C.K.W. and E.C.C.; Experimental resources and
1166 models, A.A.M, C.C., M.Bu., L.M., G.B., C.K., M.P., M.Ba.; Data Curation, B.L. and M.W.;
1167 Manuscript preparation – Original Drafting, B.L. and M.W. with additional input from all
1168 authors.

1169

1170

1171

1172

1173 **Competing Interests Statement:**

1174

1175 The authors declare no competing interests.

1176

1177

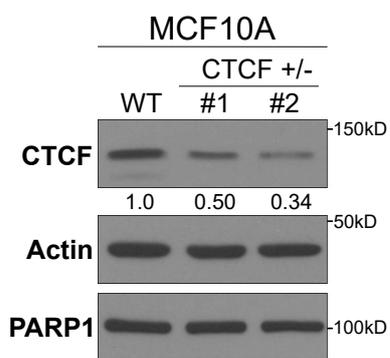
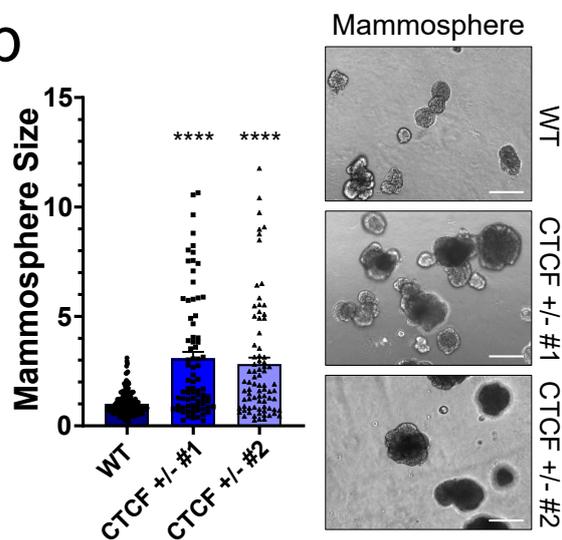
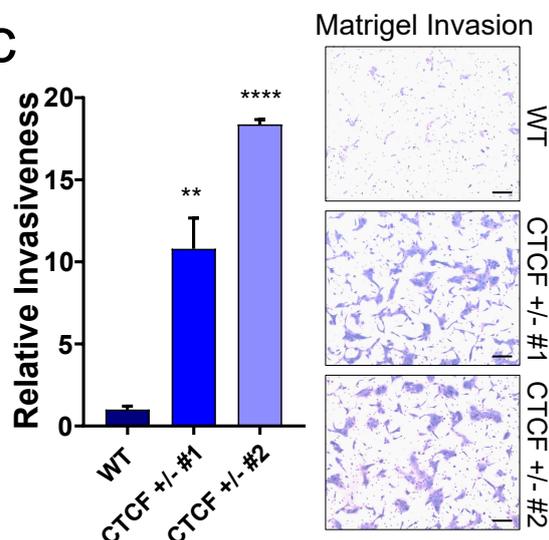
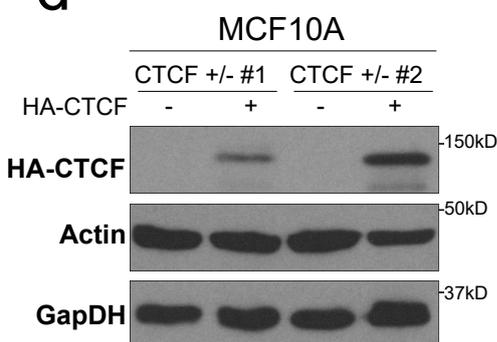
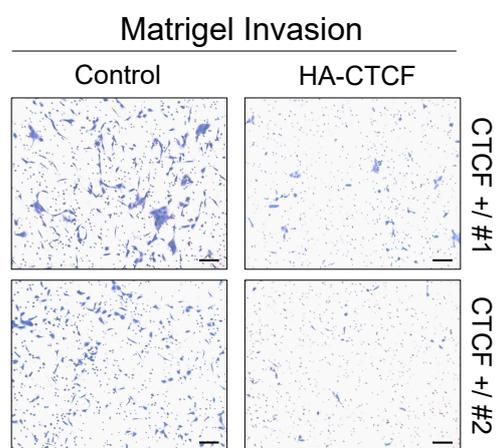
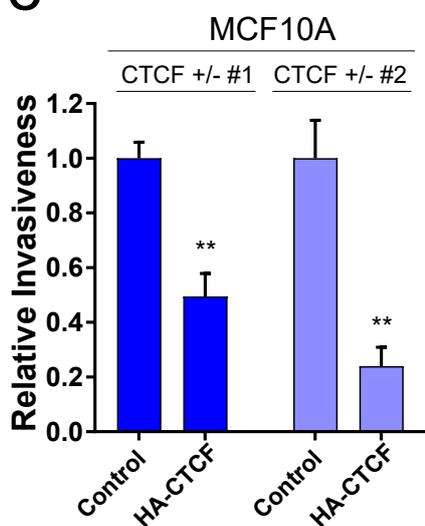
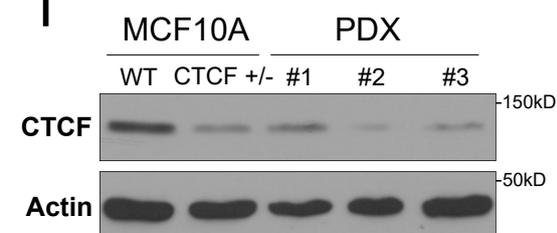
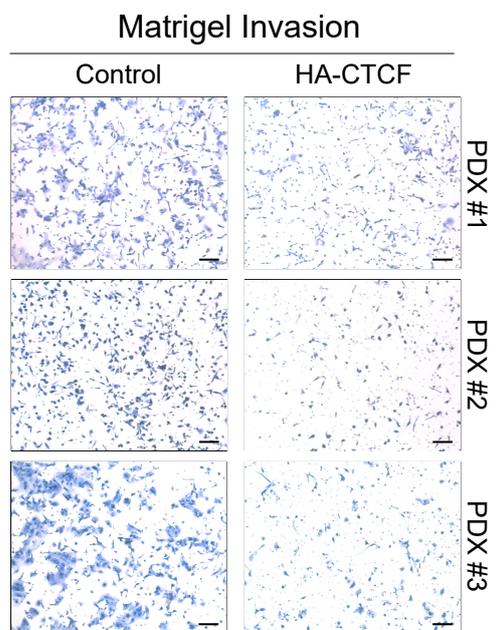
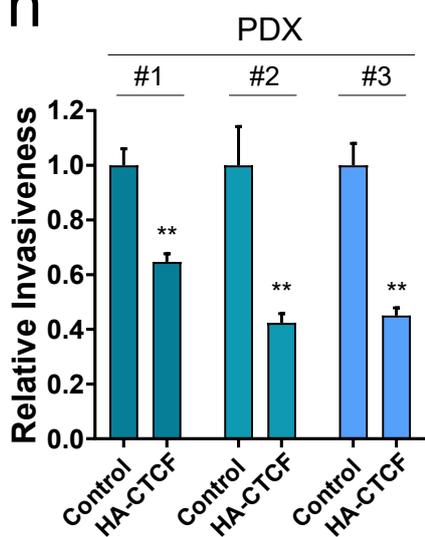
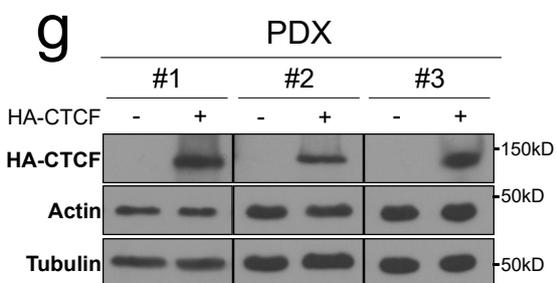
a**b****c****d****e****f****h****g**

Figure 1. CTCF LOH promotes invasiveness and unorganized growth

(a) Western Blot showing the reduction in CTCF levels in the CTCF+/- compared to WT MCF10A. Quantification represent the relative CTCF band intensity compared to WT. Actin and PARP1 are used as loading controls.

(b) Bar Chart (mean \pm SEM) representing the high average mammosphere size in the CTCF+/- cells, relative to the average WT mammosphere size. Each dot represents the size of a single mammosphere, showing that the distribution of mammosphere in the CTCF+/- is bigger and more widespread. Both p-values < 0.0001 were calculated with two-tailed Student's T Test comparing each CTCF+/- population of mammosphere to the WT. Adjacent pictures show the bigger size and deformity of the mammosphere.

(c) Bar Chart (mean \pm SEM) representing the relative invasiveness of the CTCF+/- MCF10A compared to WT. Adjacent pictures show the CTCF+/- cells able to cross the matrigel covered insert, while WT cells were mostly unable to cross. p = 0.0066 and p < 0.0001 for CTCF+/- #1 and #2, calculated using a one-tailed Student's T Test

(d) Western Blot showing the addition of HA-CTCF in the CTCF+/- cells. Actin and GapDH are used as loading control.

(e) Bar chart (mean \pm SEM) representing the loss of invasiveness of the CTCF+/- cells following the addition of HA-CTCF. p-value = 0.0013 for CTCF+/- #1 and #2, calculated using a one-tailed Student's T Test comparing HA-CTCF to control. Adjacent pictures demonstrate the nearly complete restoration of invasiveness.

(f) Western Blot showing the low basal levels of CTCF, similar to the CTCF+/- MCF10A CTCF levels, in the PDX cells. Actin is used as a loading control.

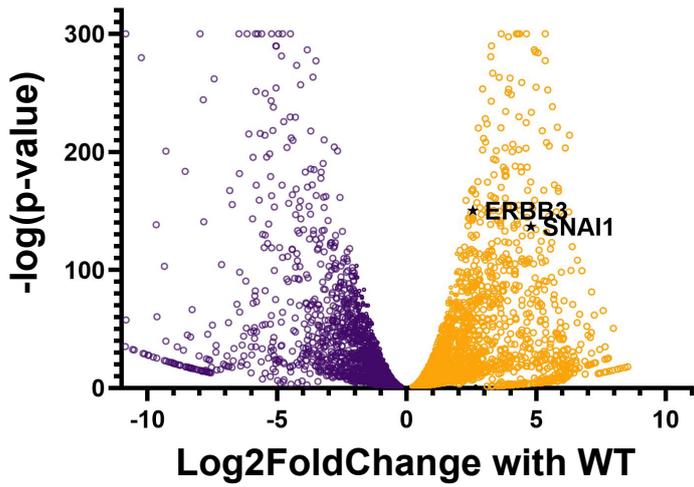
(g) Western Blot showing the addition of HA-CTCF in the 3 CTCF low PDX cell lines. Actin and Tubulin are used as loading control.

(h) Bar Chart (mean \pm SEM) representation of the decreased invasiveness in each HA-CTCF PDX cell lines relative to their own control. p-value = 0.0031, 0.0084 and 0.0015 for PDX #1, 2 and 3 respectively, calculated using a one-tailed Student's T Test comparing the HA-CTCF to control. Adjacent pictures demonstrate the noticeable reduction in invasion quantified in the bar chart.

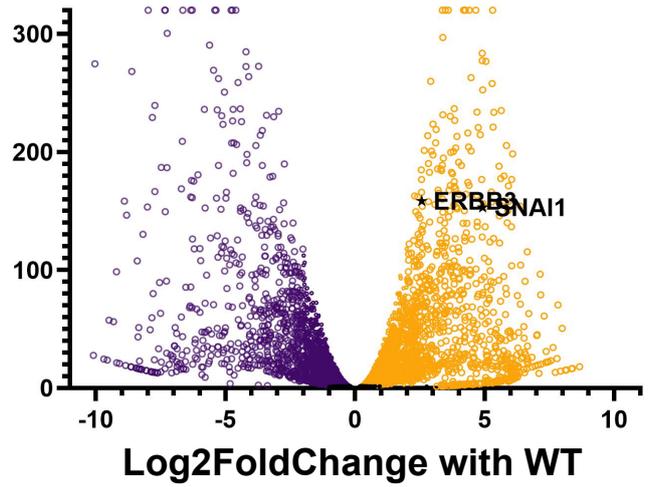
See also Supplementary Figure 1.

a

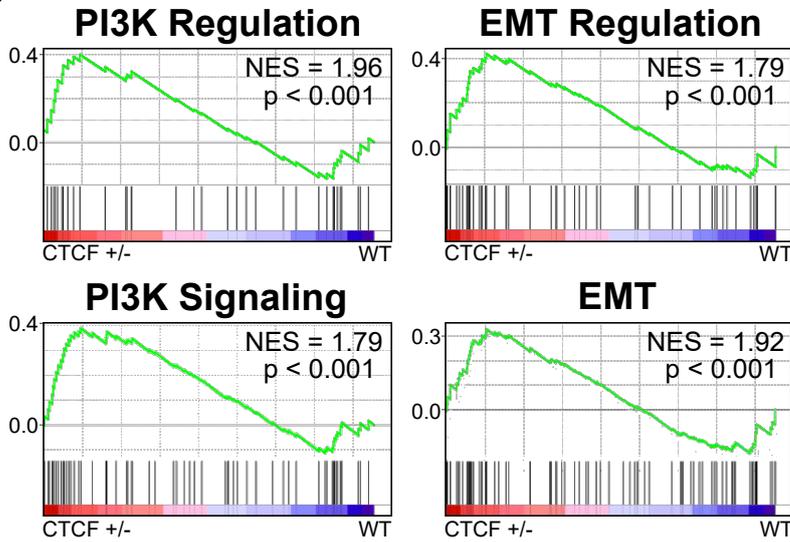
RNA-Seq CTCF +/- #1



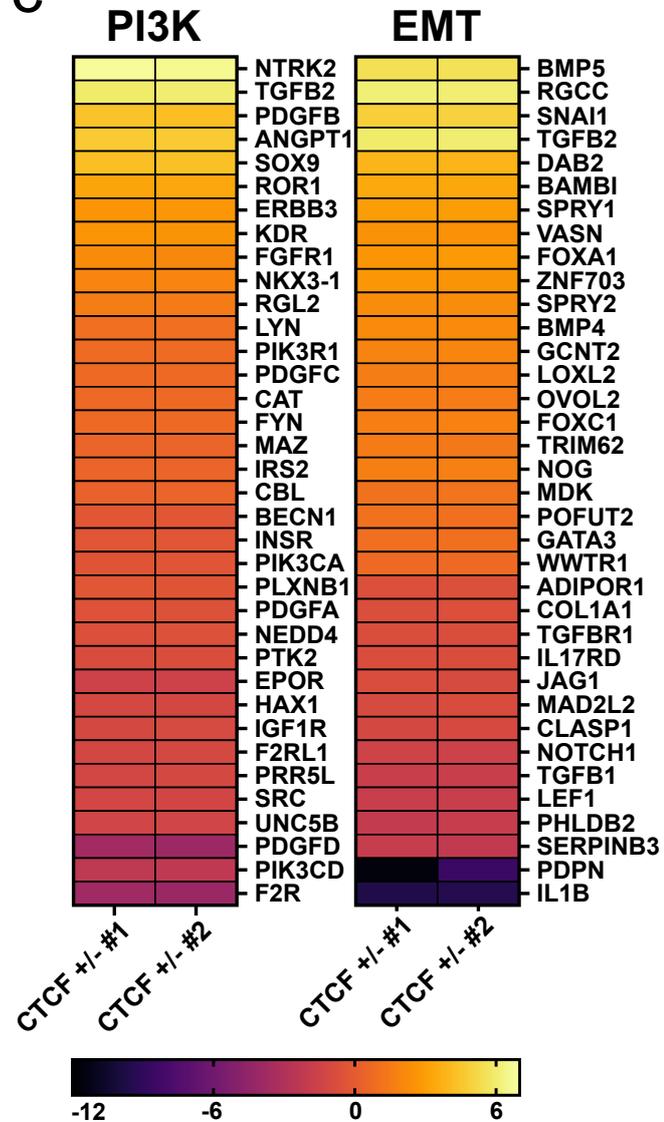
RNA-Seq CTCF +/- #2



b



c



d

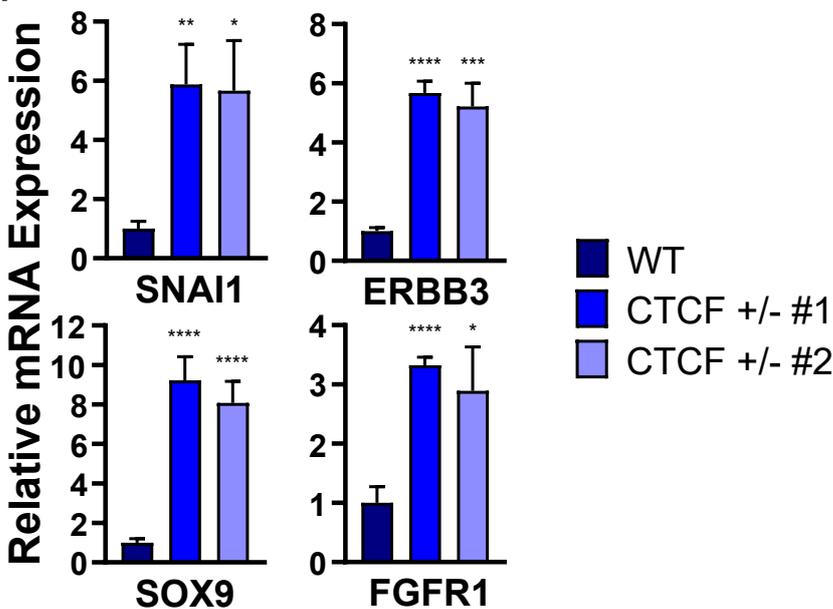


Figure 2. RNA-Seq reveals oncogenic expression underlying the invasive phenotypes

(a) Volcano Plot highlighting reproducibility of changes in gene expression between CTCF+/- . Upregulated genes are in orange, downregulated genes are in purple.

(b) GSEA Analysis highlighting the upregulation of distinct PI3K Pathway and EMT Pathway from the Gene Ontology data set in the CTCF+/- MCF10A compared to WT. Complete name of the pathways can be found in the Method section.

(c) Heatmaps of the top up and downregulated genes in PI3K and EMT Regulation pathway presented in Figure 2B.

(d) Bar chart (mean \pm SEM) of the qPCR validation of the top hits in the RNA-Seq, showing the relative expression normalized on 3 housekeeping genes compared to WT. p-values are listed below and were calculated using a two-tailed Student's T Test comparing CTCF+/- #1 and #2 to WT: SNAI1 #1 = 0.0075 and #2 = 0.0175 ; ERBB3 #1 < 0.001 and #2 = 0.0004 ; SOX9 #1 and #2 < 0.0001 ; FGFR1 #1 < 0.0001 and #2 = 0.0379.

See also Supplementary Figure 2.

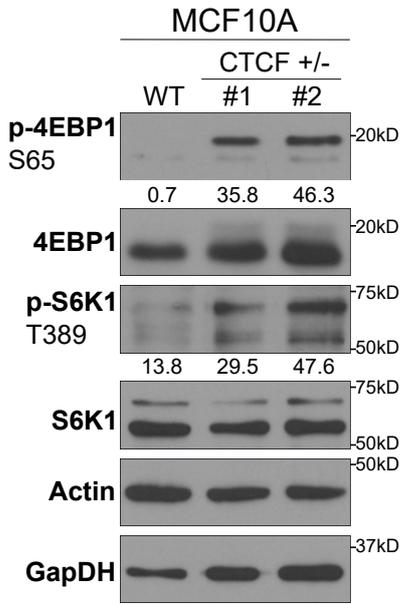
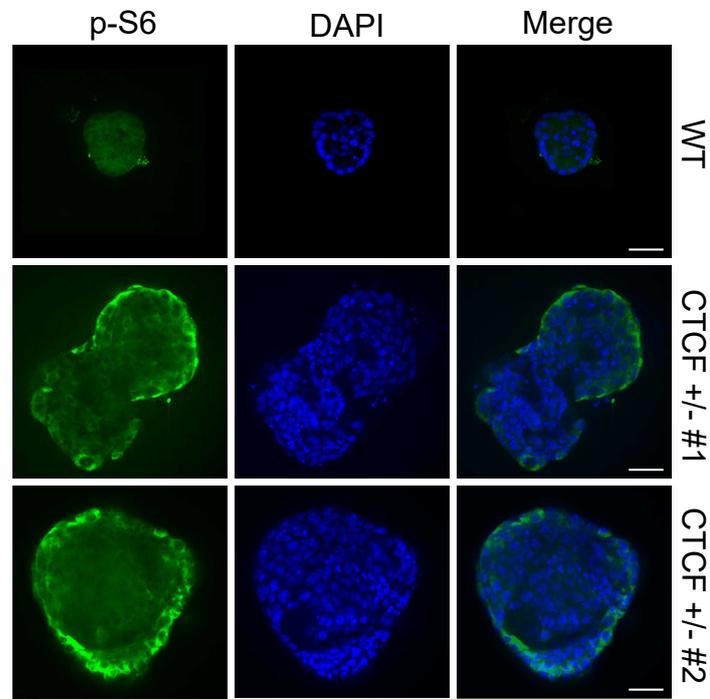
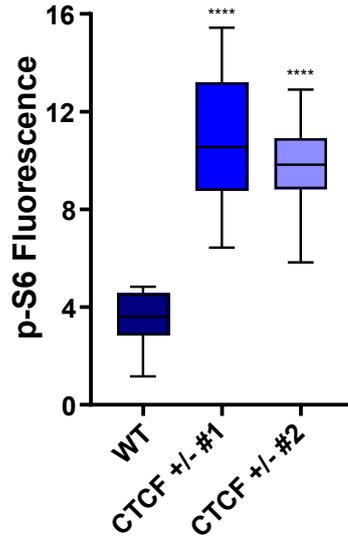
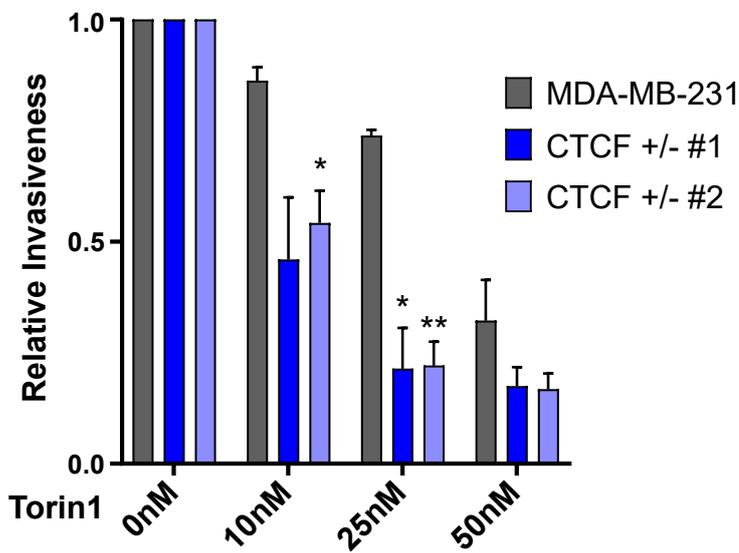
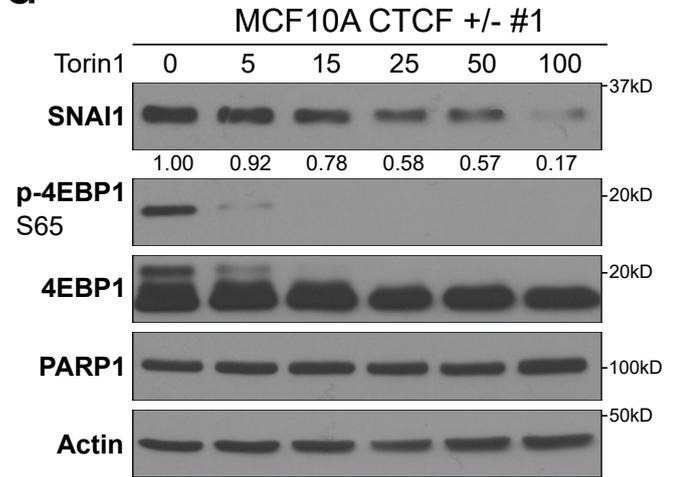
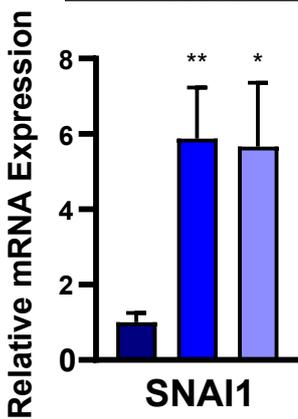
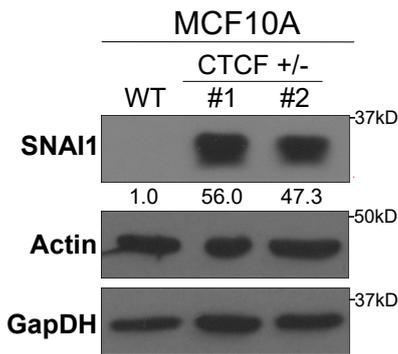
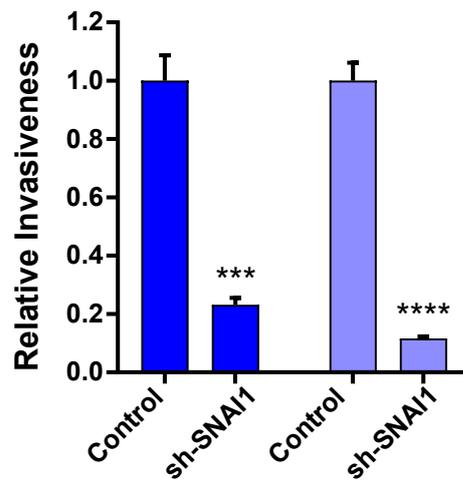
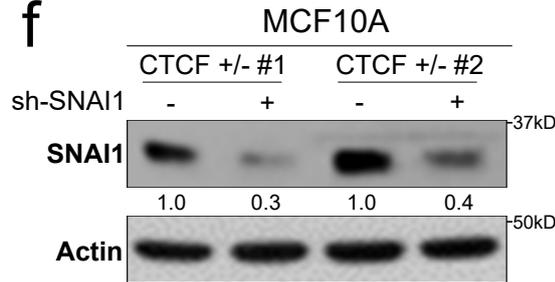
a**b****c****d****e****f**

Figure 3. The PI3K pathway and SNAI1 are synergistically driving the CTCF+/- cells oncogenicity

(a) Western Blot showing the overactivation of mTORC1 in starved condition. Quantification of the p-4EBP1 and p-S6K1 represented the band intensity normalized on background. Actin and GapDH were used as loading controls.

(b) Mammosphere immunofluorescence and quantification of S6 fluorescence of the outer layer of the mammosphere showing the overactivation of mTORC1 in the CTCF+/- mammosphere in 3D culture conditions. Represented as a Min to Max Box Plot. P-values < 0.0001 for CTCF+/- #1 and #2 were calculated using a one-tailed Student's T Test of the average intensity of S6 fluorescence in CTCF+/- mammosphere outer layer compared to WT.

(c) Bar chart (mean \pm SEM) of the quantification of invasiveness following a matrigel transwell invasion assay and Torin1 treatment. Data were normalized on each cell lines respective invasiveness in the untreated condition. P-values are listed below and were calculated with a one-tailed Student's T Test comparing the CTCF+/- relative invasiveness to MDA-MB-231 cells at specific concentration of Torin1. 10nM #1 = 0.128 and #2 = 0.0446 ; 25nM #1 = 0.0189 and #2 = 0.00308

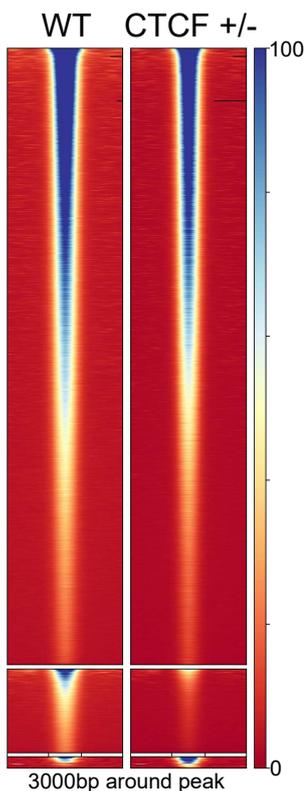
(d) Western Blot in starved condition showing the decrease of SNAI1 protein level and 4EBP1 phosphorylation following 24h Torin1 treatment. SNAI1 quantification represent the band intensity relative to untreated levels. PARP1 and Actin were used as loading controls.

(e) Western Blot showing the overexpression of SNAI1 in CTCF+/- MCF10A. SNAI1 quantification represent the band intensity relative to WT. Actin and GapDH are used as loading controls. The bar chart below represents the qPCR validation of SNAI1 overexpression at the mRNA levels shown in Figure 2D.

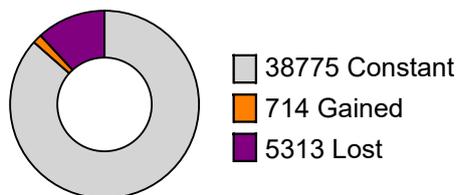
(f) Western Blot showing the reduction in SNAI1 protein levels following sh-SNAI1 treatment. SNAI1 quantification represent the band intensity relative their respective sh-CTL cell lines. Actin was used as a loading control. The bar chart (mean \pm SEM) below shows the decrease in invasiveness during a matrigel invasion assay of the sh-SNAI1 treated CTCF+/- MCF10A compared to the sh-CTL treated CTCF+/- MCF10A. p-values = 0.00103 and 0.000141 for CTCF+/- #1 and #2 respectively were calculated using a one-tailed Student's T Test comparing shSNAI1 to Control.

See also Supplementary Figure 3.

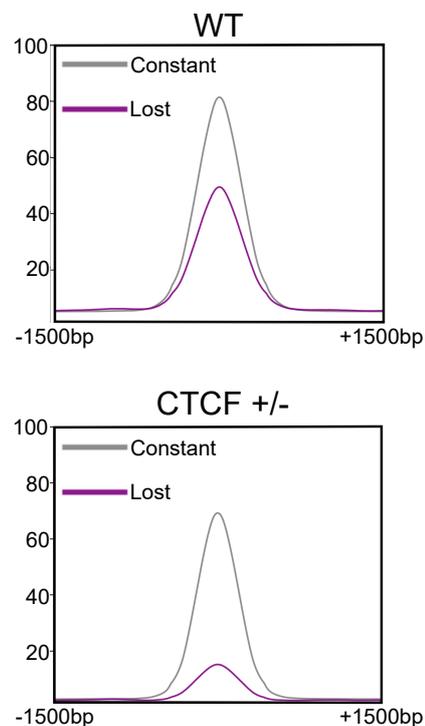
a CTCF ChIP-Seq



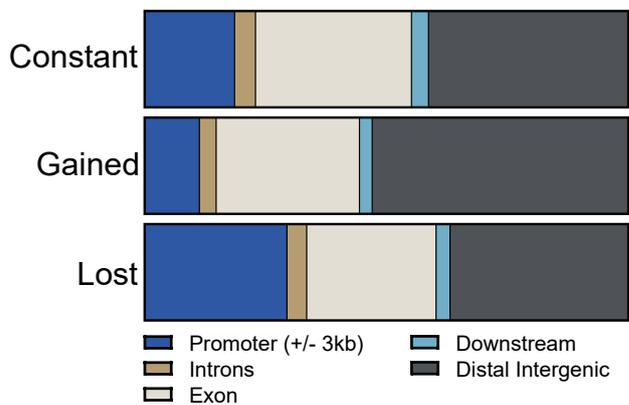
b CTCF Sites Quantification



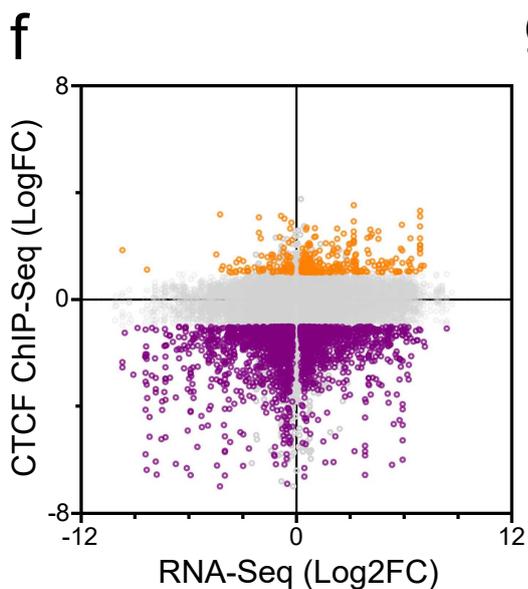
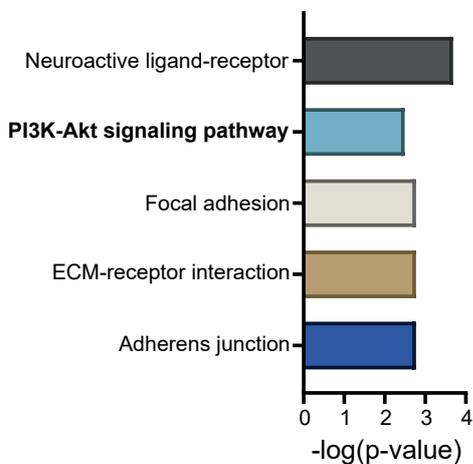
c CTCF Binding Sites



d CTCF Sites Distribution



e KEGG Enrichment at CTCF Lost Sites



g Transcription Start Sites

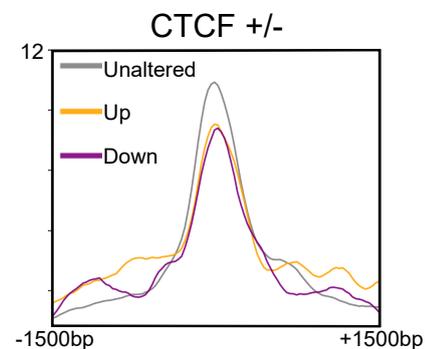


Figure 4. CTCF depletion alters CTCF DNA binding pattern

- (a) Heatmap of ChIP-Seq on CTCF, showing the clusters of constant, lost and gained sites (from top to bottom) identified by DiffBind3.0 at +/-1500bp around the sites of CTCF binding.
- (b) Pie Chart showing the number of sites in each constant, lost and gain clusters showed on the heatmap in 4A.
- (c) Profile plot of the average read density of the constant and lost cluster in the MCF10A WT and CTCF+/-, showing that the lost cluster in the WT are basally weaker binder than the constant sites and that this cluster's average read density is further reduced in the CTCF+/- cells.
- (d) Graphical representation of the relative genomic distribution of the different clusters, showing a slight enrichment of the lost sites on promoter regions.
- (e) Bar Chart showing the Top 5 enriched pathway following KEGG Pathway analysis of the lost sites of CTCF binding (ranked by geneRatio).
- (f) Dot Plot showing that lost sites of CTCF are associate with both lost and gain of gene expression. The changes in gene expression (Log2FC) between CTCF+/- and WT MCF10A are on the X axis. The changes in CTCF binding (LogFC) between CTCF+/- and WT MCF10A on the binding sites colocalizing (+/- 3kb) with each individual genes are on the Y axis. Orange dots have a CTCF logFC ≥ 1 while purple dots have a CTCF logFC ≤ 1 .
- (g) Profile plot of the average CTCF ChIP-Seq read density in CTCF+/- MCF10A at the TSS of all unaltered genes (in grey, Basemean > 100, adjusted p-value > 0.05, abs(Log2FC) < 0.5), upregulated genes (in orange, Basemean > 100, adjusted p-value < 0.05, Log2FC > 1) and downregulated genes (in purple, Basemean > 100, adjusted p-value < 0.05, Log2FC < -1) showing a slight reduction on altered genes.

See also Supplementary Figure 4.

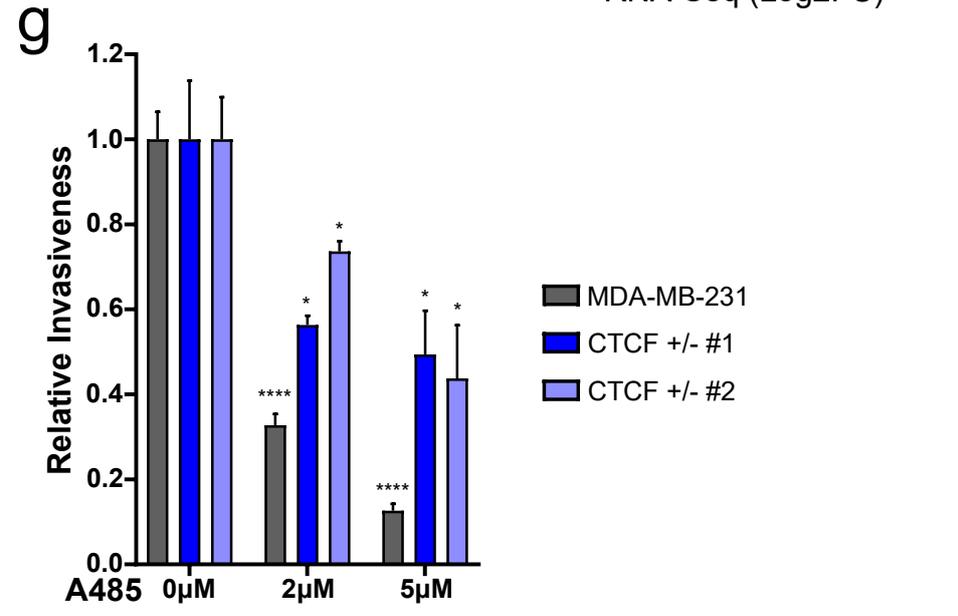
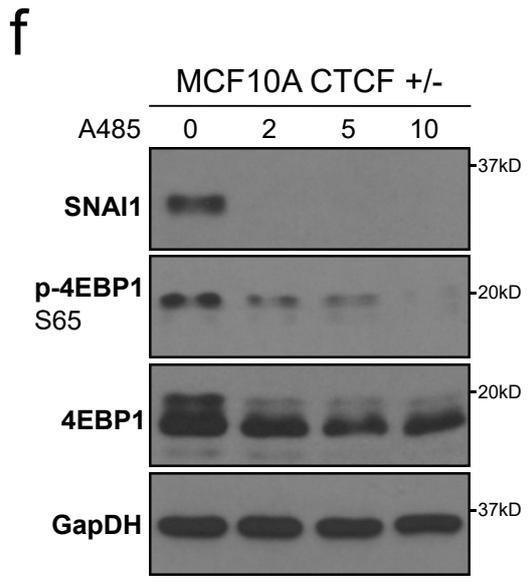
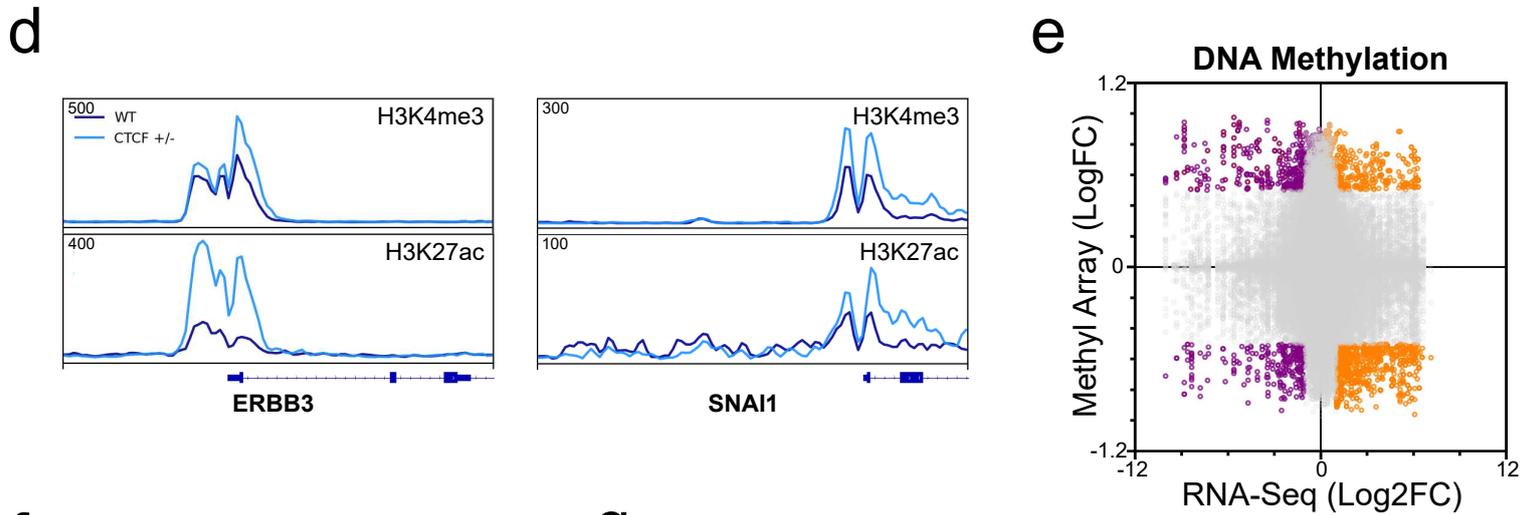
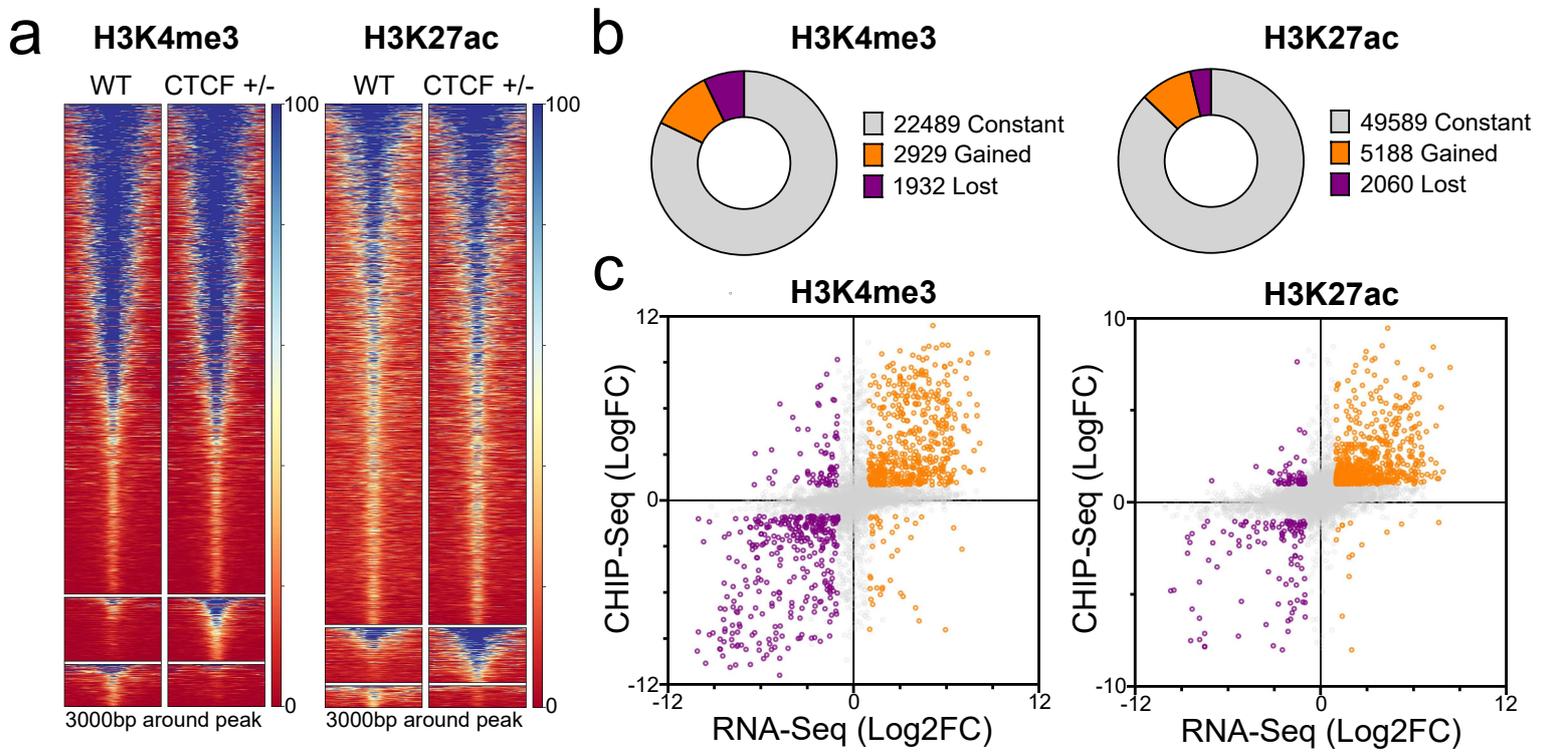


Figure 5. Epigenetic reprogramming of activating histone marks drive changes in gene expression

(a) Heatmap of ChIP-Seq on H3K4me3 and H3K27ac, showing the clusters of constant, gained and lost sites (from top to bottom) identified by DiffBind3.0 at +/-1500bp around the sites of H3K4me3 or H3K27ac binding.

(b) Pie Chart showing the number of sites in each constant, gain and lost clusters showed on the heatmaps in 5A.

(c) Dot plot correlating the changes in H3K4me3 and H3K27ac ChIP-Seq (LogFC between CTCF+/- and WT MCF10A) on the Y axis and changes in gene expression detected with the RNA-Seq (Log2FC between CTCF+/- and WT MCF10A), showing a strong and direct correlation. Orange dots have an RNA-Seq log2FC ≥ 1 while purple dots have an RNA-Seq log2FC ≤ 1 , both have an absolute CTCF LogFC ≥ 1 .

(d) ChIP-Seq track of normalized read density track showing the gain of H3K27ac and H3K4me3 on ERBB3 and SNAI1. The read density of each track ranges from 0 to the number noted in the top left corner.

(e) Dot Plot correlating the changes in DNA methylation, detected with Methyl EPIC Array (LogFC of CTCF+/- and WT MCF10A) and changes in gene expression detected with the RNA-Seq (Log2FC between CTCF+/- and WT MCF10A), showing no noticeable correlation between all points. Orange dots have an RNA-Seq log2FC ≥ 1 while purple dots have an RNA-Seq log2FC ≤ 1 , both have an absolute Methyl Array LogFC ≥ 0.5 .

(f) Western Blot in starved condition showing the reduction of 4EBP1 (Serine 65) phosphorylation and SNAI1 expression following 48h treatment with HATi A485.

(g) Bar chart (mean \pm SEM) of the relative invasiveness compared to DMSO treated of A485 treated CTCF+/- MCF10A and WT MDA-MB-231 cells showing a significant inhibitory effect of A485 in all cell lines. P-values are listed below and were calculated using a one-tailed Student's T Test comparing the invasiveness at different concentration of A485 to invasiveness of untreated cells. 2 μ M MDA < 0.0001 , #1 = 0.0444 and #2 = 0.0232; 5 μ M MDA < 0.0001 , #1 = 0.0284 and #2 = 0.0252.

See also Supplementary Figure 5.

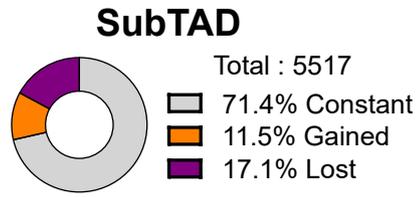
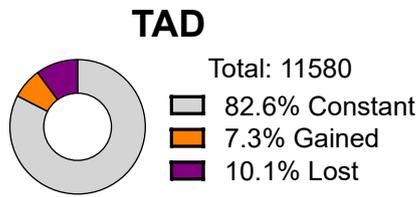
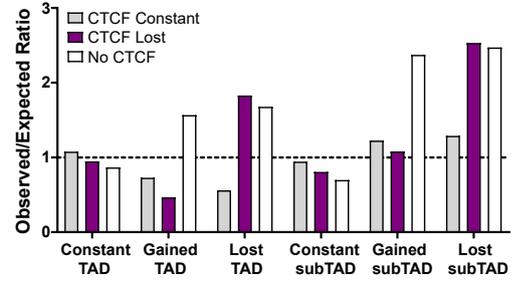
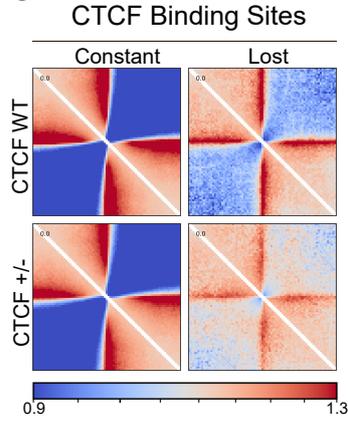
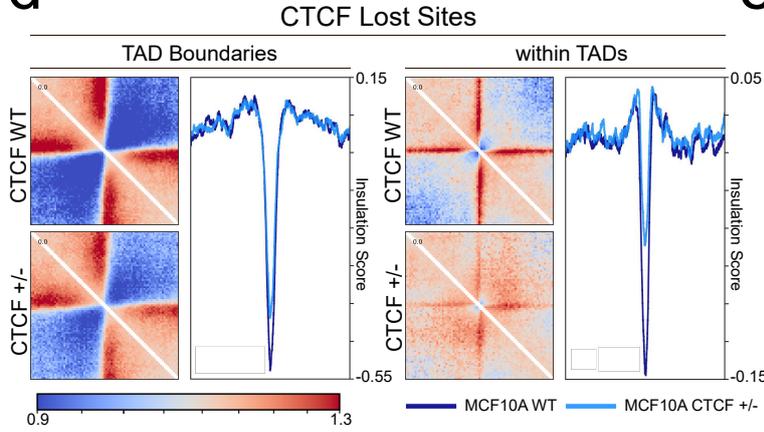
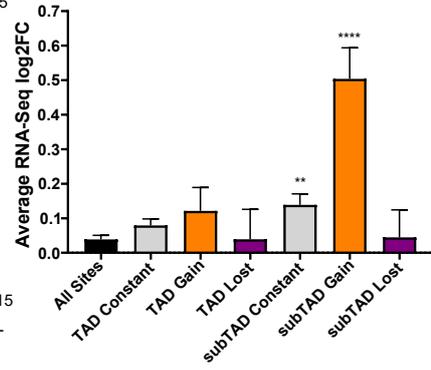
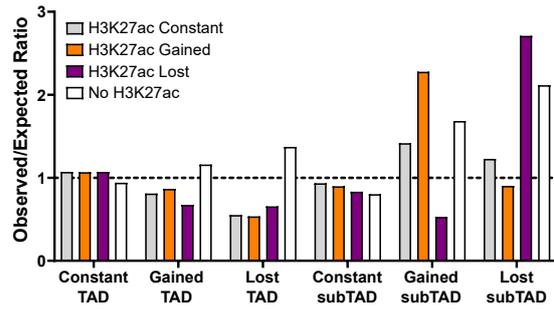
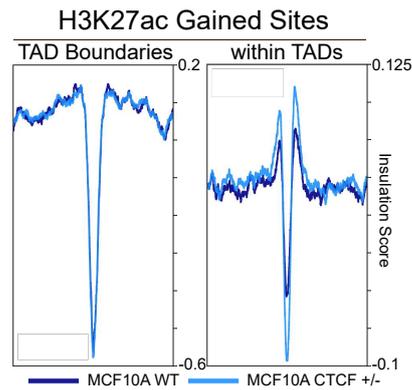
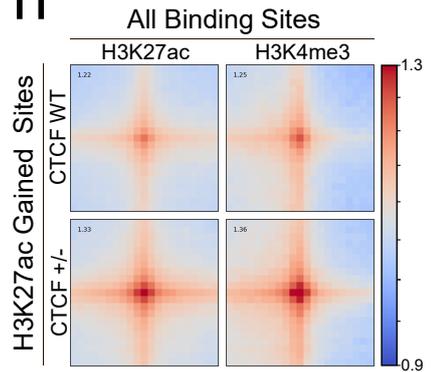
a**b****c****d****e****f****g****h**

Figure 6. Loss of subTAD insulation explains epigenetic reprogramming

- (a) Pie chart of the relative quantification of constant, gained and lost TAD and subTAD boundaries (\pm 10kb), showing more pronounced changes in subTAD boundaries.
- (b) Bar chart (O/E Ratio) demonstrating the enrichment of lost boundaries at lost sites of CTCF.
- (c) Pile-up plot showing the local interaction around constant and lost sites of CTCF (\pm 200kb), showing a loss of insulation around lost sites of CTCF. Color Scale represent the average interaction relative to randomize average genome-wide interaction.
- (d) Pile-up plot showing that subTAD insulation at CTCF lost sites is almost completely lost while TAD insulation at CTCF lost sites is slightly altered. Color Scale represent the average interaction relative to randomize average genome-wide interaction. Profile plot beside the pile-up plot represent the average insulation score of the regions showed in the pile-up plot.
- (e) Bar chart (mean \pm SEM) of the average RNA-Seq log₂FC between CTCF \pm and WT of genes colocalizing with Tad and subTAD boundaries (\pm 10kb) showing the significant transcriptional upregulation at gained subTAD boundaries.
- (f) Bar chart (O/E Ratio) showing the enrichment of gained H3K27ac on gained subTAD and vice-versa for lost H3K27ac and lost subTAD.
- (g) Profile plot of the average insulation score at sites of gained H3K27ac colocalizing with TAD boundaries or within TAD (\pm 10kb for the colocalization, \pm 200kb for the range of the profile plot), showing the gain of insulation, therefore subTAD boundary formation, at gained site of H3K27ac within TADs.
- (h) Pile-up plot showing the gain of interaction between gained H3K27ac and all sites of H3K27ac and H3K4me3 (\pm 50kb).

See also Supplementary Figure 6.

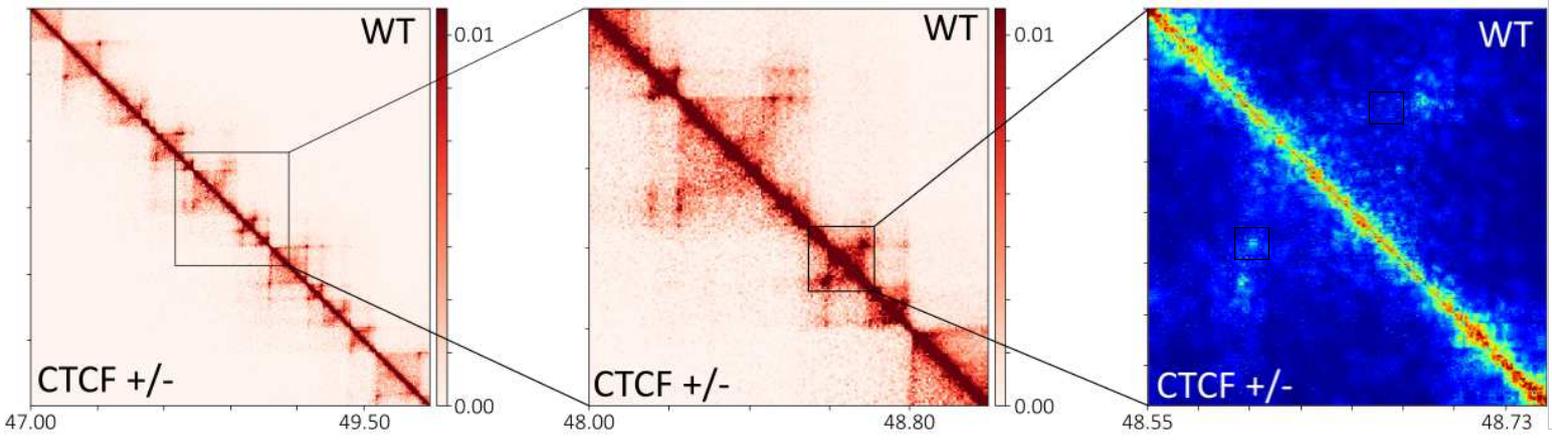
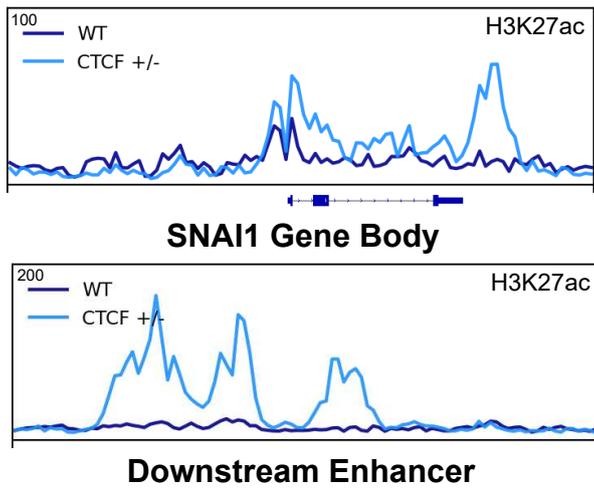
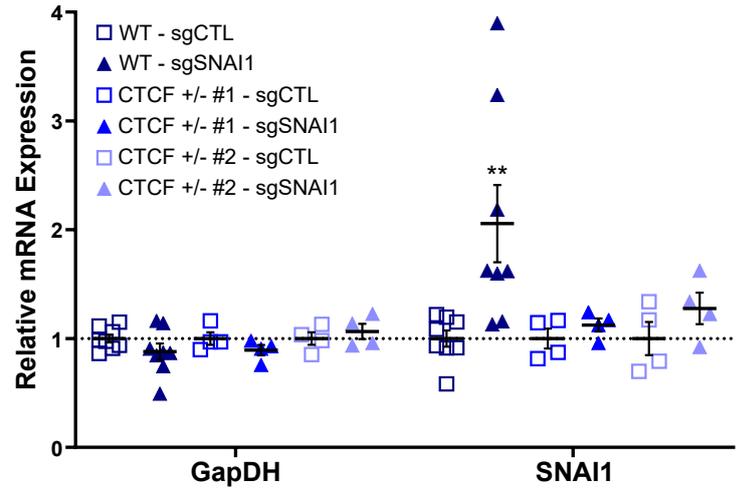
a**b****c**

Figure 7. Loss of CTCF at SNAI1 drives reorganization of subTAD interactions

(a) Gradual zoom of 10kb and 5kb resolution HiC heatmap to HIFI high-resolution heatmaps around SNAI1 loci showing the gain of enhancer promoter interaction on SNAI1 body (shown in the black box) and the absence of major conformational changes at TAD levels. Coordinates below the heatmap represent the location (in Mb) on chromosome 20.

(b) ChIP-Seq track of normalized read density track showing the gain of H3K27ac on SNAI1 gene body and the downstream enhancer which displayed a gain of interaction in figure 5J. The range of read density of each track is noted in the top left corner.

(c) Chart (mean \pm SEM) representing the mRNA expression, relative to sgCTL, of infected WT and CTCF \pm MCF10A, showing a specific increase of SNAI1 expression in sgSNAI1 WT compared to GapDH expression in sgSNAI1 WT and expression of both genes in sgSNAI1 CTCF \pm , which are unchanged. P-value = 0.0057 was calculated using Student's T Test comparing SNAI1 expression in sgSNAI1 WT and sgCTL WT. All other sgSNAI1 to sgCTL comparison were non-significant.

See also Supplementary Figure 7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CompilledSupplementalInformation.pdf](#)