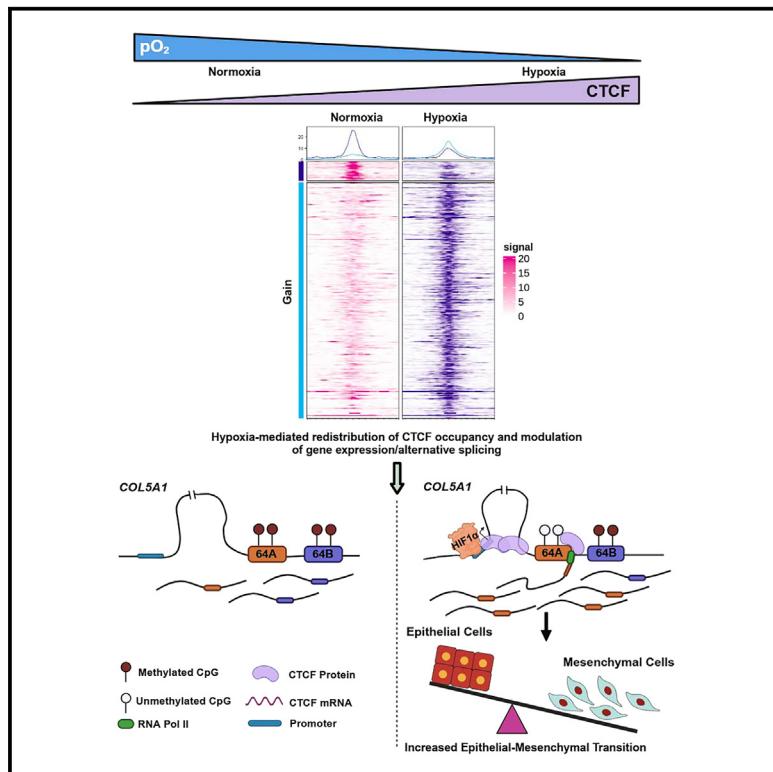


Hypoxia-induced CTCF mediates alternative splicing via coupling chromatin looping and RNA Pol II pause to promote EMT in breast cancer

Graphical abstract



Authors

Parik Kakani, Shruti Ganesh Dhamdhere, Deepak Pant, ..., Anchala Pandey, Dimple Notani, Sanjeev Shukla

Correspondence

sanjeevs@iiserb.ac.in

In brief

Kakani et al. reported that hypoxia drives CTCF redistribution across the genome and is associated with the EMT. CTCF-mediated promoter-exon upstream looping regulates the CTCF-mediated RNA polymerase II pause by modulating the DNA methylation at the distal exon and regulating the alternative splicing to favor the EMT.

Highlights

- Hypoxia drives global redistribution of CTCF occupancy across chromatin
- HIF1 α -CTCF-COL5A1exon 64A axis drives EMT in hypoxic breast cancer cells
- CTCF-driven promoter-exon upstream loop tunes DNA methylation at alternative exons
- Loss of CTCF-mediated loop switches alternative splicing and impairs EMT



Article

Hypoxia-induced CTCF mediates alternative splicing via coupling chromatin looping and RNA Pol II pause to promote EMT in breast cancer

Parik Kakani,¹ Shruti Ganesh Dhamdhere,¹ Deepak Pant,¹ Rushikesh Joshi,^{1,3} Sachin Mishra,^{2,4} Anchala Pandey,¹ Dimple Notani,² and Sanjeev Shukla^{1,5,*}

¹Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, Madhya Pradesh 462066, India

²National Center for Biological Sciences, Tata Institute for Fundamental Research, Bangalore, Karnataka 560065, India

³Present address: CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi 110025, India

⁴Present Address: Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria

⁵Lead contact

*Correspondence: sanjeevs@iiserb.ac.in

<https://doi.org/10.1016/j.celrep.2025.115267>

SUMMARY

Hypoxia influences the epithelial-mesenchymal transition (EMT) through the remodeling of the chromatin structure, epigenetics, and alternative splicing. Hypoxia drives CCCTC-binding factor (CTCF) induction through hypoxia-inducible factor 1-alpha (HIF1 α), which promotes EMT, although the underlying mechanisms remain unclear. We find that hypoxia significantly increases CTCF occupancy at various EMT-related genes. We present a CTCF-mediated intricate mechanism promoting EMT wherein CTCF binding at the collagen type V alpha 1 chain (COL5A1) promoter is crucial for COL5A1 upregulation under hypoxia. Additionally, hypoxia drives exon64A inclusion in a mutually exclusive alternative splicing event of COL5A1exon64 (exon64A/64B). Notably, CTCF mediates COL5A1 promoter-alternatively spliced exon upstream looping that regulates DNA demethylation at distal exon64A. This further regulates the CTCF-mediated RNA polymerase II pause at COL5A1exon64A, leading to its inclusion in promoting the EMT under hypoxia. Genome-wide study indicates the association of gained CTCF occupancy with the alternative splicing of many cancer-related genes, similar to the proposed model. Specifically, disrupting the HIF1 α -CTCF-COL5A1exon64A axis through the dCas9-DNMT3A system alleviates the EMT in hypoxic cancer cells and may represent a novel therapeutic target in breast cancer.

INTRODUCTION

The primary cause of breast cancer-related deaths is metastasis.^{1,2} Tumor hypoxia (a low oxygen stress condition), a key micro-environmental factor in solid tumors, activates hypoxic signaling to increase plasticity and promote the epithelial-mesenchymal transition (EMT) to drive the first step of metastasis.³⁻⁵ Tumor cells adapt to hypoxia majorly by activating the hypoxia-inducible factor (HIF) transcription factors, which modulate the expression of many genes.⁶⁻⁸ CCCTC-binding factor (CTCF), a multifunctional DNA-binding protein that is critical for regulating the epigenetic and chromatin structures of the genome, gene expression, and alternative splicing, has been shown in our earlier study to accelerate the EMT in breast cancer under hypoxia.⁹ In the previous study, we demonstrate that the tripartite axis, hypoxia-induced DNA demethylation at the CTCF promoter mediated by Tet methylcytosine dioxygenase 2 (TET2) led to the binding of HIF1 α and upregulates the CTCF expression in hypoxic breast cancer cells. This HIF1 α -CTCF axis is one important axis that favors the invasiveness of breast cancer cells.⁹

However, the mechanism underlying HIF1 α -CTCF-axis-mediated EMT under hypoxia has not been elucidated in detail. To address this, in this study, we investigated the mechanism by which CTCF regulates its target genes to favor the EMT under hypoxia. In the present study, we showed the hypoxia-driven global redistribution of CTCF occupancy, where CTCF-gained sites are related to many cancer genes. Notably, we identified collagen type V alpha 1 chain (COL5A1) as a CTCF-related, hypoxia-induced EMT gene, and hypoxia-induced CTCF regulates COL5A1 expression as well as alternative splicing to promote an invasive phenotype. We demonstrate that hypoxia-driven CTCF modulates COL5A1 promoter interactions within the gene body and promotes exon64A inclusion in COL5A1 mRNA, which favors the EMT. The inclusion of exon64A is regulated by a CTCF-mediated intricate mechanism coupling CTCF-mediated promoter-exon upstream looping with CTCF-mediated RNA polymerase II (RNA Pol II) pausing at exon64A. Finally, we use an epigenome editing strategy using the dCas9-DNMT3A system for potent, specific, and stable disruption of CTCF binding, and hence, the HIF1 α -CTCF-COL5A1exon64A axis alleviates the EMT potential of breast cancer cells under hypoxia.



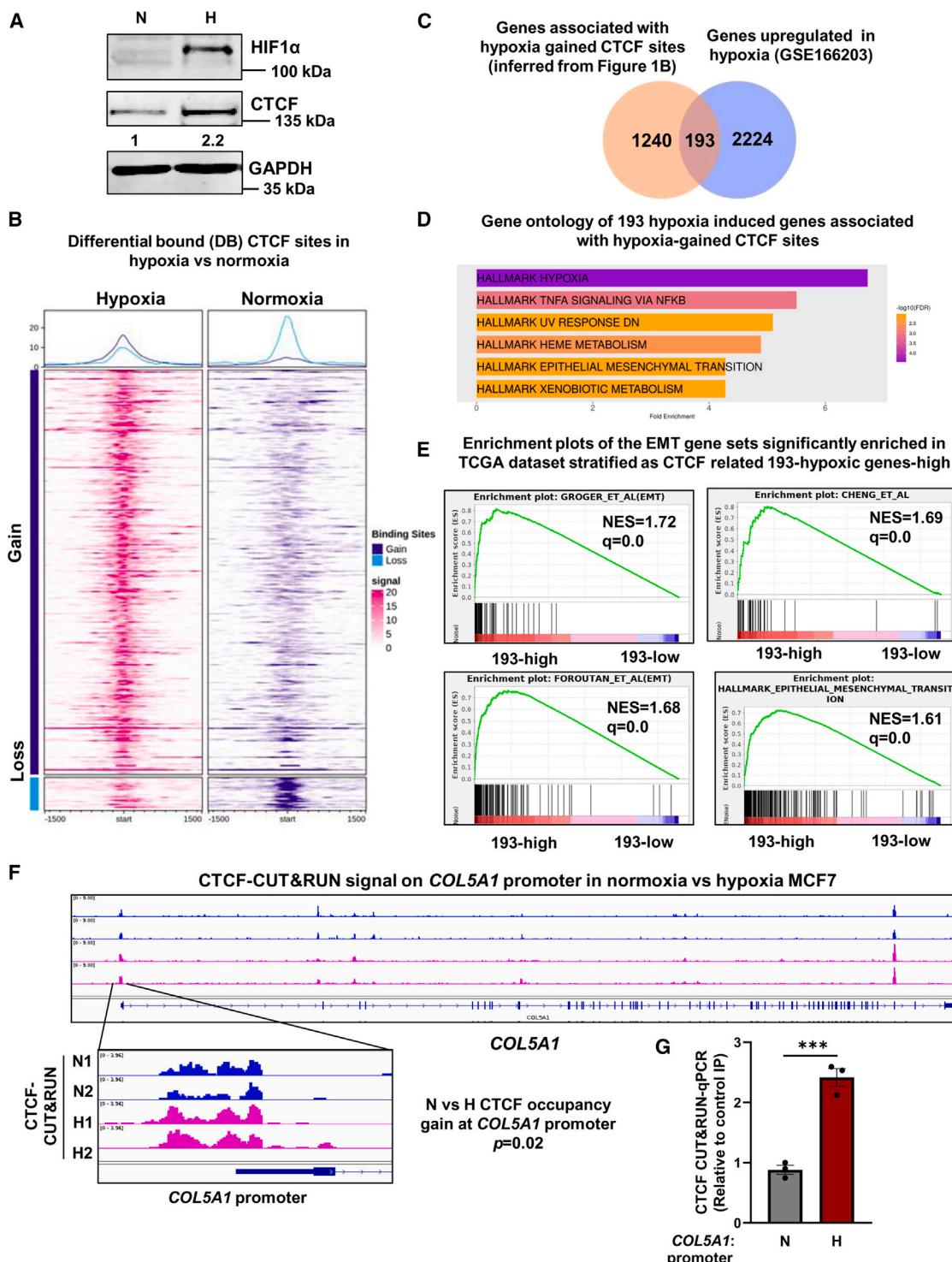


Figure 1. Hypoxia-gained CTCF occupancy regulates the gene signature associated with EMT

(A) Immunoblot of CTCF and HIF1 α in normoxic and hypoxic MCF7; GAPDH is used as loading control here (and elsewhere).

(B) Differential binding analysis of CTCF occupancy between hypoxic and normoxic MCF7 cells inferred from CTCF-CUT&RUN analysis. The top 1,000 differentially bound sites are shown.

(C) Venn diagrams showing the overlap between hypoxia-gained CTCF site-related protein-coding genes (from B) and hypoxia-upregulated genes (from GEO: GSE166203).

(D) Representative significantly enriched terms from the Gene Ontology of 193 hypoxia-induced genes associated with hypoxia-gained CTCF sites.

(legend continued on next page)

Here, we showed a hypoxia-driven, CTCF-dependent tripartite relationship between epigenetic reprogramming, alternative splicing, and the EMT in the cancer cell. Thus, our data demonstrate a specific function of the chromatin organizer CTCF for hypoxia-driven EMT by modulating alternative splicing and chromatin organization.

RESULTS

Hypoxia-gained CTCF occupancy associated with genes related to EMT

Hypoxia promotes the EMT and metastasis,^{3,4} and we found that CTCF is upregulated under hypoxic conditions in breast cancer cells (Figure 1A).⁹ This led us to hypothesize that CTCF may regulate EMT-related genes to drive a metastatic phenotype under hypoxia. To comprehensively study the genomic repertoire of CTCF binding across normoxic and hypoxic cancer cells, we performed both cleavage under targets and release using nuclease (CUT&RUN) assay and chromatin immunoprecipitation sequencing (ChIP-seq) and analyzed the differential CTCF global occupancy from normoxia to hypoxia. Our analysis of CTCF-CUT&RUN data showed a significant redistribution of CTCF occupancy where 2,156 sites were gained and 973 sites were lost from normoxia to hypoxia (Figure 1B; Table S1). Since CTCF regulates the EMT under hypoxia,⁹ we were keen to see how many CTCF gained sites were present on the promoter of EMT-related genes upregulated under hypoxia. RNA-seq data analysis showed that there are 2,472 genes upregulated under hypoxia and related to glycolysis, MTORC1 signaling, tumor necrosis factor alpha (TNF α) signaling, and the EMT (Figure S1A). Further analysis showed that 129 EMT genes are associated with hypoxia-gained CTCF occupancy on their promoter (Table S1), of which several genes were upregulated under hypoxia (Figures S1C-S1H; Table S1). Furthermore, the EMT can be regulated directly via transcription of EMT genes or indirectly by other pathways involved in the EMT,^{10,11} so we wanted to decipher whether the hypoxia-gained CTCF occupancy-related transcriptome is involved in the EMT. The overlap of CTCF gained sites with gene promoters was carried out, and we found that these 2,156 gained sites correspond to the promoter of 1,433 protein-coding genes that are associated with pathways involved in the EMT (Figure S1B; Table S1). Further, to identify the CTCF-associated transcriptome under hypoxia, we overlapped hypoxia-gained CTCF occupancy-related genes with 2,472 genes upregulated under hypoxia. We found that there are 193 hypoxia induced genes (including many EMT genes) to be associated with hypoxia-gained CTCF sites (Figure 1C). This suggests that CTCF plausibly regulating the expression of these genes under hypoxia (Figure 1C). Gene Ontology analysis of these 193 genes revealed the significant enrichment of pathways related to hypoxia, TNF α signaling via nuclear factor κ B (NF- κ B),

and also the EMT, consistent with our hypothesis that CTCF under hypoxia is plausibly associated with many EMT-related pathways (Figure 1D).

Next, to understand the association between the expression of 193 CTCF-related hypoxic genes and the prevalence of the EMT in patient samples, we used the breast cancer TCGA dataset. For this, we stratified TCGA breast dataset based on the expression of these 193 CTCF-related hypoxic genes and took the top 20% as 193 CTCF-related hypoxic gene expression high (193-high) and the bottom 20% as 193 CTCF-related hypoxic gene expression low (193-low). Further, gene set enrichment analysis (GSEA) of hallmark gene sets revealed the enrichment of multiple EMT-related datasets in TCGA samples stratified as 193-high, including hallmark EMT, transforming growth factor β (TGF- β) signaling, Groger et al.¹² (EMT), and Foroutan et al.¹³ (EMT) (Figure 1E; Table S1). When running a leading edge analysis with the EMT gene sets significantly enriched in 193-high, *COL5A1* and *WNT5B* were most represented in the enrichment score of the significant gene sets (Figure S1I). Intriguingly, we further observed higher increased expression of *COL5A1* than *WNT5B* in hypoxia (Table S1). Moreover, we observed hypoxia-gained CTCF occupancy near the transcriptional start site of *COL5A1* in comparison to the normoxic condition (Figures 1F, 1G, and S2A).

Recent studies have highlighted the involvement of collagens in cancer progression,¹⁴⁻¹⁷ but only a limited number of studies have investigated their functional role in cancer metastasis and their contribution to hypoxia-driven cancer progression. Therefore, *COL5A1*, being (1) a hallmark gene of both EMT and hypoxia (MSigDB) and (2) a hypoxia-gained CTCF occupancy-related EMT gene most represented in the EMT dataset enriched in TCGA dataset stratified as 193-high and (3) given its novelty as a direct CTCF target and its unexplored role in the EMT in hypoxic conditions, was chosen as a model gene to study the transcriptional reprogramming of EMT genes and the invasiveness mediated by CTCF in hypoxia-induced settings.

CTCF promotes expression of a hypoxia-induced EMT gene, *COL5A1*

We next determined the clinical significance of *COL5A1* in patients with breast cancer and observed higher *COL5A1* expression in breast cancer tumor samples than normal samples in TCGA dataset (Figure S2B). Breast cancer samples stratified as hypoxic_high had significantly higher *COL5A1* levels than those stratified as hypoxic_low (Figure S2C). We found similar results when comparing normoxic and hypoxic HCC1806 cell lines (Figure S2D). These three independent cohorts suggested that *COL5A1* may have a role in hypoxic breast cancer progression.

Further, the *COL5A1* mRNA and protein levels were increased in MCF7 and HCC1806 under hypoxic conditions (Figures 2A,

(E) Enrichment plots of the EMT gene sets significantly enriched in 193 CTCF-related hypoxic gene expression-high phenotype (enrichment score [ES] > 1) after running a GSEA on TCGA BRCA dataset, stratified as 193-high ($n = 220$) or 193-low ($n = 220$) based on the expression of 193 CTCF-related hypoxic genes.

(F) CTCF binding profile from CTCF-CUT&RUN for *COL5A1* gene is shown (top) and zoomed-in promoter region of *COL5A1* (bottom, enrichment $p = 0.02$, inferred from DiffBind) in MCF7 cells under normoxia and hypoxia.

(G) CTCF-CUT&RUN-qPCR on *COL5A1* promoter in normoxic and hypoxic MCF7 cells. Error bars, mean \pm SEM; two-tailed t test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$; $n = 3$ biological replicates.

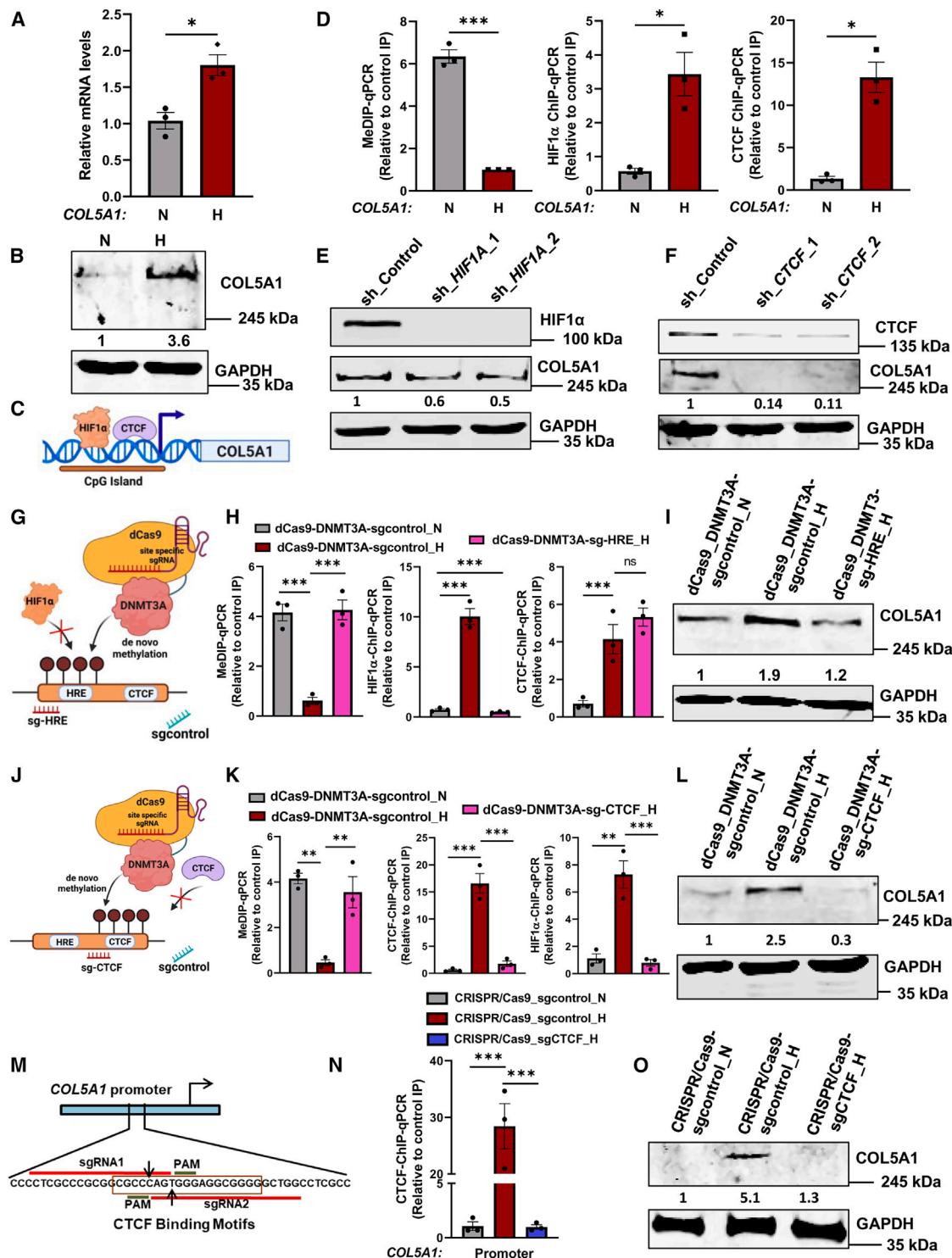


Figure 2. COL5A1 expression is mechanistically driven by the interplay between DNA methylation, CTCF, and HIF1 α in hypoxic breast cancer cells

(A and B) Expression analysis of COL5A1 using (A) RT-qPCR (normalized to RPS16 mRNA levels) and (B) immunoblot in normoxic and hypoxic MCF7 cell line.

(C) Schematic showing putative presence of HIF1 α and CTCF binding motifs on COL5A1 promoter.

(D) MeDIP-qPCR, HIF1 α , and CTCF-ChIP-qPCR on COL5A1 promoter in normoxic and hypoxic MCF7 cells.

(E) Immunoblot of HIF1 α and COL5A1 in HIF1A knockdown and shControl MCF7 cells under hypoxia.

(F) Immunoblot of CTCF and COL5A1 in CTCF knockdown and shControl MCF7 cells under hypoxia.

(legend continued on next page)

2B, S2E, and S2F). To further decode the hypoxia-mediated induction of *COL5A1*, we examined the *COL5A1* promoter and found the hypoxia responsive element (HRE) motif and CTCF binding site (inferred from CTCF peak present upstream of the transcription start site [TSS] of *COL5A1* in normoxia vs. hypoxia CTCF-CUT&RUN; Figure 1G; Table S2), further suggesting the involvement of these proteins in the regulation of *COL5A1* expression (Figure 2C). Promoter analysis also revealed the presence of dense CpG islands spanning HIF1 α and CTCF binding sites (Figure 2C), and it is reported that the binding of both CTCF and HIF1 α to genomic loci is inhibited by DNA methylation.^{18–20} Further, methylated DNA IP (MeDIP)-qPCR analysis revealed highly methylated DNA around the TSS of *COL5A1* under normoxia, which is reduced under hypoxia (Figures 2D and S2G). Furthermore, we also observed an increased enrichment of both HIF1 α and CTCF on the *COL5A1* promoter under hypoxia compared to normoxia (Figures 2D and S2G). Moreover, depleting HIF1 α or CTCF greatly diminished hypoxia-induced *COL5A1* expression (Figures 2E, 2F, S2H, and S2I).

***COL5A1* expression is mechanistically driven by the interplay between DNA methylation, CTCF, and HIF1 α in hypoxic breast cancer cells**

To investigate whether HIF1 α directly binds to the *COL5A1* promoter or reduced *COL5A1* expression in HIF1 α -depleted cells is due to its effect on CTCF expression, we performed a luciferase assay. The results demonstrated that a mutation in the HRE motif in the *COL5A1* promoter deletion construct reduced the luciferase activity by 50%; however, a mutation in either only the CTCF binding site or both the CTCF and HRE motifs did not show induction in luciferase activity under hypoxia in comparison to the normoxia control (Figure S2J). These findings indicate that HRE and CTCF motifs present within the promoter of *COL5A1* are required for the efficient activation of *COL5A1* under hypoxia.

To further validate this observation, we employed dCas9-DNMT3A with single-guide RNA (sgRNA) particularly targeting the HRE motif on the *COL5A1* promoter (Figure 2G) and observed that targeting dCas9-DNMT3A-sgRNA at the HRE motif on the *COL5A1* promoter efficiently maintained DNA methylation (with no potential off-target activity [Figure S2K]) (Figure 2H), abolished HIF1 α enrichment on the *COL5A1* promoter, and, hence, reduced *COL5A1* expression under hypoxia (Figure 2I). However, disrupting HIF1 α binding did not have any effect on hypoxia-mediated CTCF enrichment on the *COL5A1* promoter (Figure 2H). Strikingly, targeting dCas9-DNMT3A-sgRNA to the CTCF motif on the *COL5A1* promoter to maintain DNA methylation (Figure 2J) (with no potential off-target activity [Figure S2L]) resulted in

reduced enrichment of CTCF on the *COL5A1* promoter (Figure 2K) and also inhibited *COL5A1* upregulation under hypoxia (Figure 2L). In conclusion, impeding CTCF occupancy has a more substantial negative effect on *COL5A1* expression under hypoxia than inhibiting HIF1 α binding utilizing the dCas9-DNMT3A system (Figures 2I and 2L). Intriguingly, the methylation-dependent hindrance of CTCF enrichment abolished HIF1 α enrichment on the *COL5A1* promoter under hypoxia (Figure 2K). Conclusively, the upregulation of *COL5A1* is initiated by hypoxia-driven CTCF enrichment on the *COL5A1* promoter, followed by HIF1 α occupancy for its efficient induction.

Further disruption of the CTCF motif in the *COL5A1* promoter (Figures 2M and S2M) abolished CTCF occupancy in edited cells (Figure 2N) and impeded *COL5A1* induction under hypoxia (Figure 2O). Taking these results together, we demonstrate that hypoxia-driven *COL5A1* expression is collectively mediated by a tripartite mechanism involving epigenetic regulation, HIF1 α , and CTCF.

Hypoxia drives the alternative splicing event in *COL5A1* and favors exon64A inclusion in breast cancer cells

Previous studies have identified a mutually exclusive alternative splicing event of exon64 of the *COL5A1* gene that gives rise to an isoform that contains either exon64A or exon64B^{21,22} (Figure 3A). To investigate whether hypoxia drives the alternative splicing of exon64, we first analyzed the alternative splicing event of exon64 in a breast cancer patient dataset and found that exon64A is significantly included in samples stratified as hypoxic_high (Figure 3B). Similarly, HCC1806 HTA 2.0 array data revealed the significant inclusion of exon64A under hypoxia (Figure 3C). Further, qPCR analysis showed the significant inclusion of exon64A and the exclusion of exon64B under hypoxia in breast cancer cell lines (Figures 3D and S3A). These findings indicated that hypoxia-mediated alternative splicing of exon64 and hypoxic breast cancer cells specifically prefers the inclusion of exon64A over exon64B. Therefore, we sought to further define the molecular mechanisms driving *COL5A1* exon64 alternative splicing under hypoxia.

Hypoxia-driven CTCF occupancy regulates the inclusion of exon64A in *COL5A1* mRNA and promotes EMT in breast cancer cells

It is noteworthy to mention that CTCF is important for exon inclusion by pausing RNA Pol II-mediated transcription in a methylation-dependent manner.^{18,19,23} Interestingly, our analysis of CTCF-CUT&RUN assay and CTCF-ChIP-seq revealed hypoxia-gained CTCF occupancy on exon64A in comparison to the

(G) Schematic representation of targeting the HRE motif present in the *COL5A1* promoter region by dCas9-DNMT3A with specific sgRNA against HRE to maintain methylation and suppress HIF1 α occupancy.

(H and I) MeDIP-qPCR, HIF1 α , and CTCF-ChIP-qPCR on *COL5A1* promoter (H) and immunoblot of *COL5A1* (I) in MCF7 cells transfected with dCas9-DNMT3A-sgHRE vs. sgcontrol under hypoxia in comparison to the normoxia sgcontrol.

(J) Schematic showing specific sgRNA against CTCF motif to maintain methylation and suppress CTCF enrichment on *COL5A1* promoter using dCas9-DNMT3A.

(K and L) MeDIP-qPCR, CTCF, and HIF1 α -ChIP-qPCR on *COL5A1* promoter (K) and immunoblot of *COL5A1* (L) in MCF7 cells transfected with dCas9-DNMT3A-sgCTCF vs. dCas9-DNMT3A-sgcontrol under hypoxia in comparison to the normoxia sgcontrol.

(M) Schematic showing sgRNA sequence targeting the CTCF motif present in the *COL5A1* promoter using CRISPR-Cas9.

(N and O) CTCF-ChIP-qPCR on *COL5A1* promoter (N) and immunoblot of *COL5A1* (O) in CRISPR-Cas9-sgCTCF hypoxic cells in comparison to the normoxic and hypoxic sgcontrol MCF7 cells.

Error bars, mean \pm SEM; two-tailed t test; * p < 0.05, ** p < 0.01, and *** p < 0.001; n = 3 biological replicates.

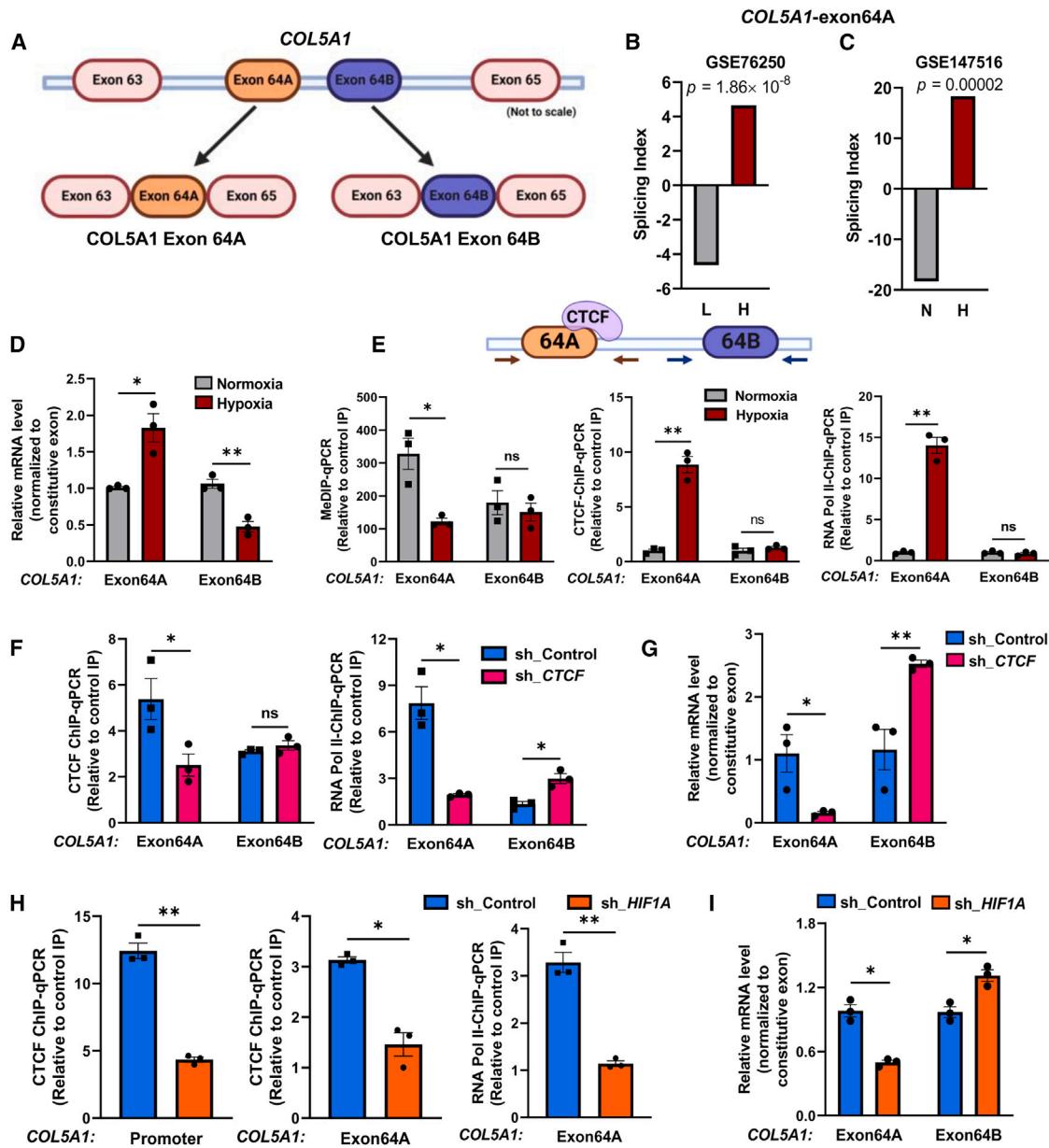


Figure 3. Hypoxia-induced COL5A1 exon64A inclusion is regulated by DNA demethylation-dependent enrichment of CTCF on exon64A and is associated with RNA Pol II pausing

(A) COL5A1 gene exon64 has two mutually exclusive isoforms, exon64A and exon64B.

(B) Significant inclusion of exon64A of COL5A1 in GEO: GSE76250 breast cancer dataset stratified as hypoxic_high (H, $n = 41$) or hypoxic_low (L, $n = 41$).

(C) Significant inclusion of exon64A of COL5A1 in microarray data in HCC1806 (GEO: GSE147516) post-hypoxic treatment ($n = 2$) and represented by splicing index values.

(D) RT-qPCR analysis of COL5A1 exon64A and exon64B isoforms normalized to RPS16 and constitutive exon expression levels in normoxic and hypoxic MCF7 cells.

(E) Schematic showing putative CTCF binding site on exon64A, MeDIP-qPCR, CTCF, and RNA Pol II ChIP-qPCR on exon64A and exon64B in normoxic and hypoxic MCF7 cells.

(F and G) CTCF and RNA Pol II ChIP-qPCR on exon64A and exon64B (F) and RT-qPCR analysis of COL5A1 exon64A and exon64B isoforms normalized to RPS16 and constitutive exon expression levels (G) in MCF7 cells transduced with either shCTCF or shControl under hypoxia.

(H and I) CTCF-ChIP-qPCR on COL5A1 promoter and CTCF and RNA Pol II ChIP-qPCR on COL5A1 exon64A (H) and RT-qPCR analysis of COL5A1 exon64A and exon64B isoforms normalized to RPS16 and constitutive exon expression levels (I) in MCF7 cells transduced with either shHIF1A or shControl under hypoxia. Error bars, mean \pm SEM; two-tailed t test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$; $n = 3$ biological replicates.

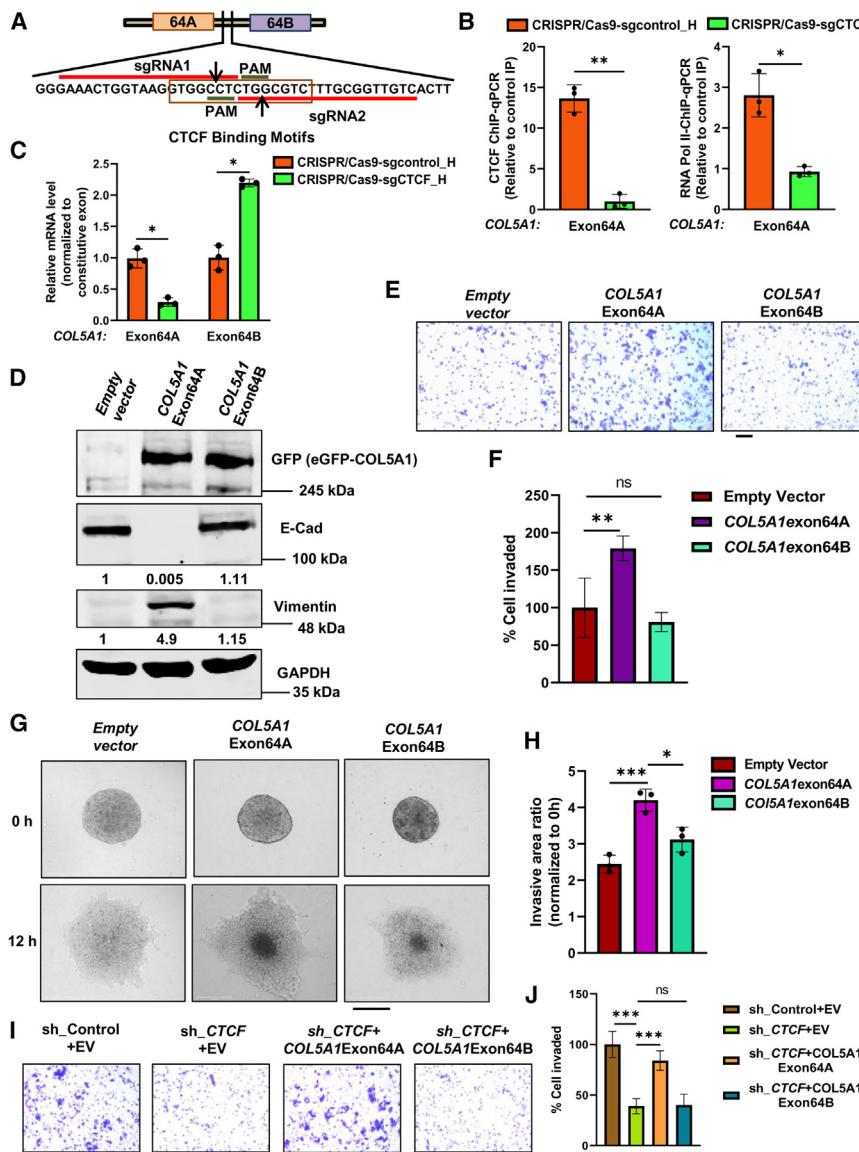


Figure 4. Hypoxia-driven, CTCF-mediated inclusion of exon64A in COL5A1 mRNA promotes EMT in breast cancer cells

(A) Schematic showing sgRNA sequence targeting the CTCF motif present immediately downstream of exon64A.

(B and C) CTCF and RNA Pol II ChIP-qPCR on COL5A1 exon64A (B) and RT-qPCR analysis of COL5A1 exon64A and exon64B isoforms normalized to RPS16 and constitutive exon expression levels (C) in MCF7 cells transduced with CRISPR-Cas9-sgRNAs specific against CTCF motif immediately downstream of exon64A in comparison to the sgcontrol MCF7 cells under hypoxia.

(D) Immunoblot of GFP (to confirm overexpression of COL5A1 isoforms), E-cad, and vimentin in MCF7 cells with empty vector (EV), COL5A1exon64A, or COL5A1exon64B ectopically expressed in normoxic MCF7 cells.

(E and F) Invasion assay (scale bars, 200 μ m) (E) with its quantification as a percentage of cells invaded (F) after overexpression of EV, COL5A1exon64A, or COL5A1exon64B in MCF7 cells under hypoxia.

(G and H) Representative images (scale bars, 275 μ m) (G) with their quantification (H) of 3D spheroid invasion assay in MCF7 cells overexpressing EV, COL5A1exon64A, or COL5A1exon64B spheroids under normoxic conditions.

(I and J) Invasion assay (scale bars, 200 μ m) (I) with its quantification (J) after overexpression of EV, COL5A1exon64A, or COL5A1exon64B isoforms in CTCF-depleted MCF7 cells in comparison to the shControl cells under hypoxia.

Error bars, mean \pm SEM; two-tailed t test; * p < 0.05, ** p < 0.01, and *** p < 0.001; n = 3 biological replicates.

normoxia, and a CTCF motif search revealed the presence of a CTCF binding motif immediately downstream of exon64A (Figures 3E and S3B; Table S2), corroborating with the “road-block” model where CTCF-mediated RNA Pol II elongation stalling occurs downstream of included exons.¹⁹ Therefore, we further investigated the role of CTCF in regulating hypoxia-mediated alternative splicing of COL5A1 and the inclusion of exon64A. First, we checked the DNA methylation status of exon64 and revealed the presence of DNA methylation on both exons under normoxia; however, only exon64A showed hypoxia-dependent reduced DNA methylation (Figures 3E and S3C). Further, CTCF-ChIP-qPCR analysis revealed CTCF enrichment on exon64A and not on exon64B under hypoxia as compared to the normoxia (Figures 3E and S3C). We then validated the RNA Pol II pause and found significant RNA Pol II occupancy on exon64A under hypoxia (Figures 3E and S3C).

Notably, CTCF depletion resulted in reduced enrichment of CTCF (Figures 3F and S3D) and RNA Pol II (Figures 3F and S3D) on exon64A with a comparable shift in RNA Pol II enrichment (Figures 3F and S3D) on exon64B that resulted in reduced inclusion of exon64A, with an increased inclusion of exon64B under hypoxia (Figures 3G and S3E). This demonstrates the role of CTCF in mediating the inclusion of exon64A. Furthermore, we observed reduced CTCF and RNA Pol II enrichment on exon64A (Figure 3H), resulting in reduced inclusion of exon64A and increased inclusion of exon64B (Figure 3I) in HIF1 α -depleted cells. Collectively, COL5A1 exon64A inclusion is regulated by hypoxia-driven, DNA demethylation-dependent, CTCF-mediated RNA Pol II pausing.

Hypoxia-driven, CTCF-mediated inclusion of exon64A in COL5A1 mRNA promotes EMT in breast cancer cells

We further investigated the association between hypoxia-driven, CTCF-mediated inclusion of exon64A and the EMT phenotype of hypoxic breast cancer cells. The disruption of the CTCF motif downstream of exon64A (Figures 4A and S4A) caused reduced

occupancy of CTCF and, hence, reduced RNA Pol II on exon64A (**Figure 4B**). We further observed reduced inclusion of exon64A and increased inclusion of exon64B under hypoxia in CTCF motif edited cells as compared to the hypoxic control cells (**Figure 4C**). This strengthens our observation that hypoxia-mediated CTCF enrichment downstream of exon64A is required for the inclusion of exon64A and that disruption of this CTCF binding is enough to switch the alternative splicing outcome in the hypoxic cells. Moreover, the inclusion of exon64B is not dependent on CTCF.

We further determine the hypoxia-induced, CTCF-mediated alternative splicing event contribution in the EMT process and observed that overexpression of the *COL5A1*exon64A isoform is associated with increased expression of vimentin and reduced E-cadherin (E-cad) levels in comparison to the control cells under normoxic conditions (**Figure 4D**). However, overexpression of the *COL5A1*exon64B isoform did not show any change in the expression of EMT molecules (**Figure 4D**). Furthermore, we found that overexpression of the *COL5A1*exon64A, but not the *COL5A1*exon64B, isoform resulted in increased invasive potential of hypoxic cells (**Figures 4E** and **4F**). Likewise, the *COL5A1*exon64A isoform, not the *COL5A1*exon64B isoform, has an increased cell migration property as compared to the control cells (**Figures S4B** and **S4C**). Additionally, *COL5A1*exon64A did not show any significant change in cell proliferation; however, *COL5A1*exon64B showed significantly increased cell proliferation in comparison to the control cells (**Figure S4D**). Furthermore, a three-dimensional (3D) invasion assay revealed that spheroids from cells overexpressed with the *COL5A1*exon64A isoform, but not the *COL5A1*exon64B isoform, invaded more into the collagen matrix compared to the control cells (**Figures 4G** and **4H**). Hence, we conclude that the inclusion of *COL5A1*exon64A plays a crucial role in enhancing the invasive and migratory phenotype of breast cancer cells under hypoxic conditions. Conversely, *COL5A1*exon64B does not significantly affect the EMT phenotype; however, it does reflect a proliferative phenotype in breast cancer cells. Moreover, the restoration of *COL5A1*exon64A expression in CTCF-depleted breast cancer cell lines reversed the invasive-inhibiting effects of CTCF knockdown, but overexpression of *COL5A1*exon64B in CTCF-depleted cells did not have any effect on the invasive phenotype of CTCF-depleted cells in comparison to the control cells under hypoxia (**Figures 4I**, **4J**, **S4E**, and **S4F**). Therefore, CTCF-mediated *COL5A1*exon64A inclusion promotes the EMT under hypoxia. Accordingly, this demonstrated that hypoxia-induced CTCF expression modulates gene expression and alternative splicing events to drive the EMT phenotype of breast cancer cells.

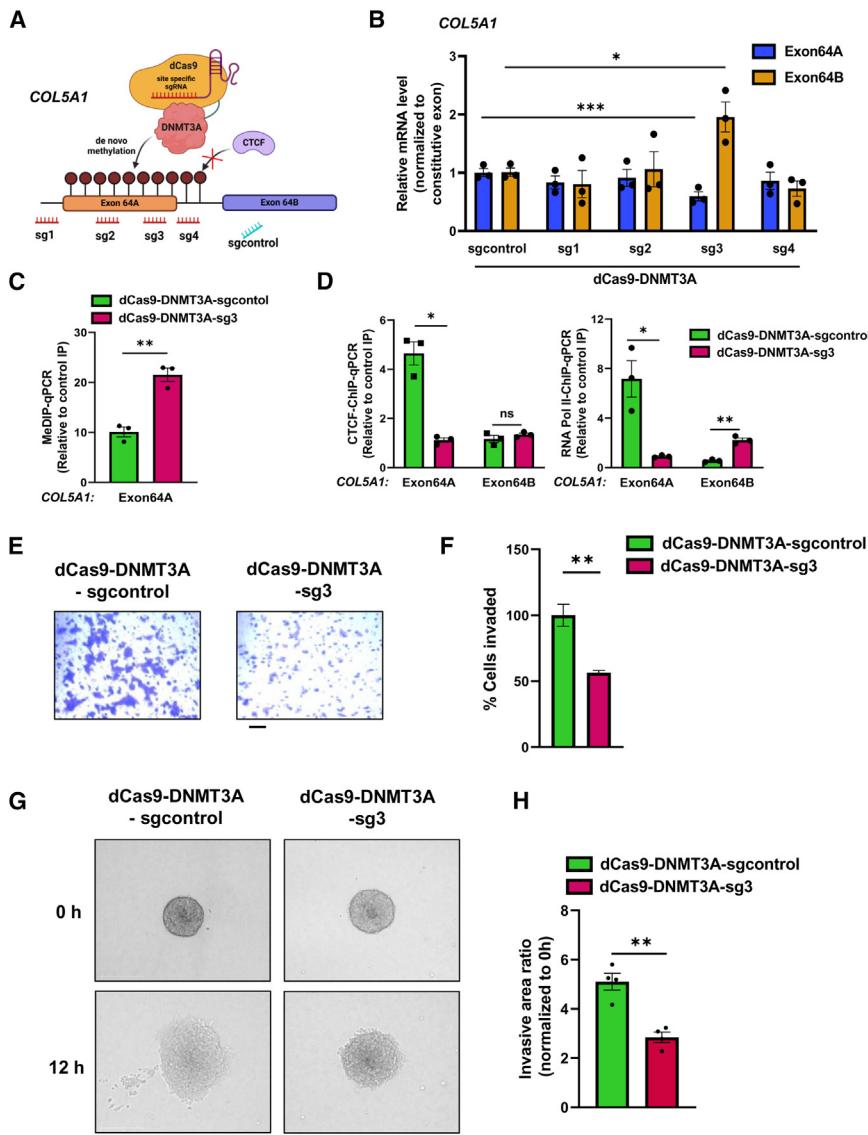
Targeted de novo methylation of CTCF binding site at exon64A alters CTCF-mediated *COL5A1* exon64 alternative splicing

Next, we assessed whether DNA demethylation and CTCF-mediated RNA Pol II pausing are sufficient to drive the inclusion of exon64A under hypoxia and employed the dCas9-DNMT3A system guided by different sgRNAs to target the CTCF binding site near exon64A of *COL5A1* (**Figures 5A** and **S5A**). Further, we observed a significant switch in the splicing event outcome of exon64, which led to reduced inclusion of exon64A and increased

inclusion of exon64B in dCas9-DNMT3A-sg3 as compared to dCas9-DNMT3A-sgcontrol under hypoxia (**Figure 5B**). MeDIP-qPCR analysis further confirmed increased methylation on exon64A with no potential off-target activity in dCas9-DNMT3A-sg3 as compared to dCas9-DNMT3A-sgcontrol (**Figures 5C** and **S5B**), which in turn abolished CTCF-mediated RNA Pol II pausing and, hence, the inclusion of exon64A under hypoxia (**Figure 5D**). Next, we observed increased wound closure in dCas9-DNMT3A-sgcontrol hypoxic cells in comparison to the normoxic cells, whereas the recruitment of dCas9-DNMT3A-sg3 to the CTCF binding site downstream of *COL5A1* exon64A resulted in reduced wound closure in comparison to the hypoxic dCas9-DNMT3A-sgcontrol cells (**Figures S5C** and **S5D**). Moreover, cells with dCas9-DNMT3A-sg3 showed increased proliferation in comparison to dCas9-DNMT3A-sgcontrol under hypoxia (**Figure S5E**), corroborating with the results that dCas9-DNMT3A-sg3 resulted in increased inclusion of the *COL5A1*exon64B isoform (**Figure 5B**), which is proliferative in nature (**Figure S4D**). Likewise, we observed a reduced invasive phenotype of dCas9-DNMT3A-sg3 cells in comparison to dCas9-DNMT3A-sgcontrol under hypoxia (**Figures 5E** and **5F**). Additionally, a 3D invasion assay demonstrated significant suppression of invasion into the collagen matrix of dCas9-DNMT3A-sg3 cells in comparison to the dCas9-DNMT3A-sgcontrol cells under hypoxia (**Figures 5G** and **5H**). These data demonstrate that hypoxia-mediated DNA demethylation and CTCF-mediated RNA Pol II pausing are sufficient to drive the alternative splicing of exon64 of *COL5A1*, promoting exon64A inclusion and, hence, the EMT phenotype under hypoxia.

CTCF-mediated promoter-exon upstream looping regulates CTCF-mediated RNA Pol II pausing and the inclusion of *COL5A1* exon64A under hypoxia

It is alluring to find that the *COL5A1* gene is 205 kb long, with the presence of an alternative splicing event for the third-to-last exon situated 188 kb downstream of the promoter. Also, CTCF-mediated promoter-proximal exon loop formation drives the alternative splicing of exons situated far from the promoter.²⁴ Thus, we assessed whether CTCF binding at the promoter and near exon64A plays any role in bringing exon64 into close physical proximity with its promoter and regulating exon64 alternative splicing under hypoxia. The circular chromosome conformation capture sequencing (4C-seq) analysis provided a comprehensive, unbiased comparison of interactions from the *COL5A1* promoter (viewpoint) across the gene loci (**Figures 6A** and **S6A**). In normoxic cells, we observed many long-range contact frequencies from the *COL5A1* promoter to other regions within the gene body and outside; however, these long-range contact frequencies decreased, and many short-range contact frequencies increased across the length of the *COL5A1* gene body (**Figures 6A** and **S6A**). These findings are largely in line with the findings from recent studies performed by the Bickmore lab,²⁵ where they demonstrated that transcription activation reduces long-range contact and induces short-range contact, corroborating with the increased transcription of the *COL5A1* gene under hypoxia. CTCF-mediated chromatin loops preferentially form between two convergently bound CTCF molecules with biases of the sense CTCF motif at the promoter or anchor site.^{26,27} The



CTCF motif at the *COL5A1* promoter (Figure 2D) is in a sense orientation, and further analysis showed three CTCF motifs in antisense orientations in the proximity of exon64 (Table S2).

Interestingly, we observed significant interactions of the *COL5A1* promoter with the T1 site only in hypoxic cells (CTCF motif situated ~7 kb upstream of exon64A) and not with the T2 (CTCF motif situated ~5.6 kb upstream of exon64A) or T3 (CTCF motif situated ~700 bp downstream of exon64A) sites in 4C-seq data located upstream of exon64A as compared to the normoxic cells (Figures 6A and S6A). Moreover, this hypoxia-specific interaction of the promoter and T1 site is lost upon depletion of CTCF under hypoxia (Figures 6A and S6A). In agreement with interactions detected in 4C-seq, we observed strong interactions between the promoter and T1 site, whereas no interactions were detected with either T2 or T3 (Figures 6B, 6C, and S6B) under hypoxia as compared to the normoxia in 3C-qPCR. Additionally, the 3C-PCR product sequencing result corroborates with the 4C-seq

Figure 5. Targeted methylation of CTCF binding site on exon64A by dCas9-DNMT3A blocks CTCF enrichment and leads to reduced inclusion of exon64A in *COL5A1* mRNA and EMT under hypoxia

(A) Schematic representation of targeting the putative CTCF binding sites immediately downstream of *COL5A1* exon64A by dCas9-DNMT3A with specific sgRNAs to maintain methylation under hypoxia to abolish CTCF enrichment.

(B) RT-qPCR of *COL5A1* exon64A and exon64B in MCF7 cells transfected with dCas9-DNMT3A-sgRNAs or sgcontrol under hypoxia.

(C and D) MeDIP-qPCR (C), CTCF-ChIP-qPCR and RNA Pol II ChIP-qPCR (D) in MCF7 cells transfected with either dCas9-DNMT3A-sgcontrol or dCas9-DNMT3A-sg3 under hypoxic condition.

(E and F) Invasion assay (scale bars, 200 μ m) (E) with its quantification (F) in MCF7 cells transfected with either dCas9-DNMT3A-sgcontrol or dCas9-DNMT3A-sg3 under hypoxic condition.

(G and H) Representative images (scale bars, 275 μ m) with their quantification (H) of 3D spheroid invasion assay in MCF7 cells transfected with either dCas9-DNMT3A-sgcontrol or dCas9-DNMT3A-sg3 under hypoxic condition.

Error bars, mean \pm SEM; two-tailed t test; * p < 0.05, ** p < 0.01, and *** p < 0.001; n = 3 biological replicates.

data (Figures S6B and S6C). Further, we observe fewer interactions between the promoter and T1 site in CTCF-depleted hypoxic cells (Figures 6C and S6B). Consistently, we found enriched CTCF occupancy only at T1 CTCF site under hypoxia (Figure 6D). Conclusively, our analysis unequivocally depicts that CTCF participates in the long-range chromatin interaction that is ~181 kb long and brings together the *COL5A1* promoter and exon64A in close proximity via intragenic CTCF looping (Figure 6B).

To define the correlation between CTCF-mediated promoter-intragenic looping and alternative splicing events in hypoxia, we designed a specific sgRNA (Figure 6E) targeting dCas9-DNMT3A to the CTCF_T1 site to investigate whether *de novo* methylation would interfere with the looping function of CTCF. We observed increased DNA methylation with no potential off-target activity (Figures 6F and S6D) and decreased CTCF enrichment at CTCF_T1 in dCas9-DNMT3A-CTCF_T1 in comparison to dCas9-DNMT3A-sgcontrol under hypoxia (Figure 6F). Interestingly, we found reduced interactions between the promoter and T1 site in dCas9-DNMT3A-sgCTCF_T1 in comparison to dCas9-DNMT3A-sgcontrol under hypoxia (Figure 6G). Further, we observed increased DNA methylation and reduced CTCF-mediated RNA Pol II pausing on exon64A and decreased inclusion of exon64A in dCas9-DNMT3A-sgCTCF_T1 in comparison to dCas9-DNMT3A-sgcontrol under hypoxia (Figure 6H). Collectively, we propose a model where hypoxia-driven *COL5A1*

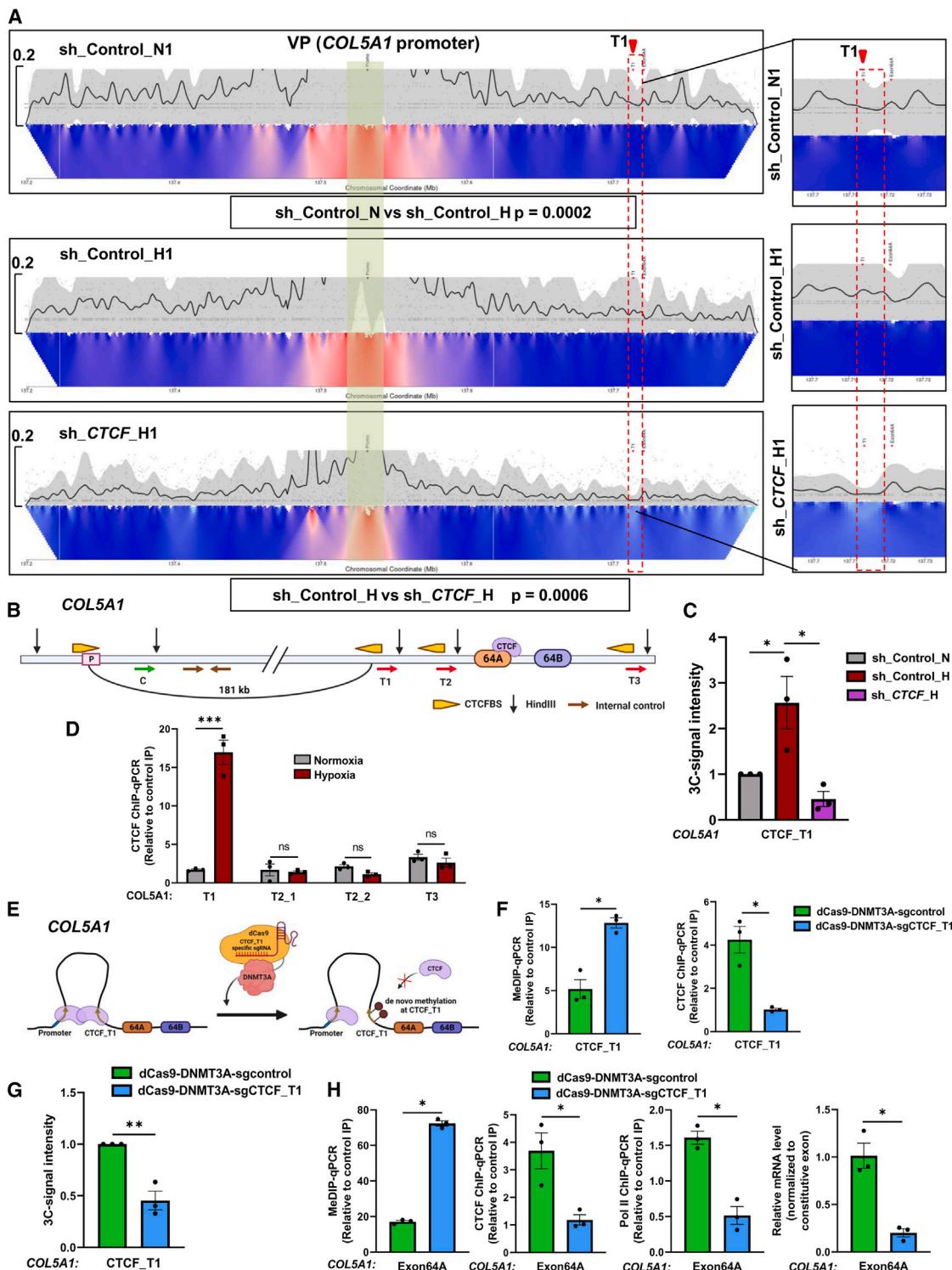


Figure 6. Hypoxia-induced CTCF-mediated COL5A1 promoter-alternatively spliced exon upstream looping decides inclusion of COL5A1 exon64A by regulating CTCF-mediated RNA Pol II pausing at exon64 under hypoxia

(A) View of genomic regions around COL5A1 with 4C-seq data. The green region is VP (viewpoint; COL5A1 promoter), and the red dotted line box is the T1 site. p values were calculated using a Student's t test based on the normalized counts in the red dotted line box, T1. The red arrows in the domainograms point to the T1 site, which shows contacts with the promoter under hypoxia as compared to the normoxia, which is lost upon CTCF depletion under hypoxia ($n = 2$). 4Cseqpipe result of a representative biological replicate.

(legend continued on next page)

alternative splicing is regulated by an intricate mechanism wherein CTCF-mediated promoter-alternatively spliced exon upstream looping modulates DNA demethylation levels of distant alternatively spliced exon64A and regulates CTCF-mediated RNA Pol II pausing at exon64A to favor its inclusion.

Hypoxia-driven CTCF differential binding regulates transcriptional and alternative splicing reprogramming in hypoxic breast cancer cells

To investigate the impending mechanisms of CTCF regulation of transcription and alternative splicing in hypoxic conditions, we performed CTCF-ChIP-seq in normoxic and hypoxic MCF7 cells. We identified 4,142 CTCF-occupied sites that were significantly gained from normoxia to hypoxia (Figure 7A; Table S3). Further analysis identified that promoters of 2,515 genes overlapped with gained CTCF peaks, suggesting direct regulation of these genes by hypoxia-driven gained CTCF occupancy (Table S3). Further analysis showed that these genes were enriched for processes related to pathways such as mTORC1, β -catenin, and TNF α -related pathways (Figure 7B). To determine the effect of CTCF-gained occupancy on hypoxic cell transcriptome diversity, we intersected the genes upregulated in hypoxic conditions with CTCF peaks significantly gained in hypoxia and identified 456 genes (Figure 7C) that were enriched in genes involved in metabolism, EMT, and hypoxia (Figure 7D).

We next investigated the relationship between significantly gained CTCF sites and alternative splicing events occurring in hypoxic cells. Further analysis showed 18,826 exon inclusion events in hypoxic conditions (Figure S7A) and 1,257 exon inclusion events pertaining to all three models of CTCF-mediated alternative splicing (Figures 7E and S7B–S7F) and corresponding to 1,137 protein-coding genes enriched in various biological processes, such as chromosome organization, the cell cycle, and RNA processing (Table S3).

We also found that genes such as *AOPEP*, *BCAS3*, *CUX1*, *RNF150*, and *ITPR1* are perhaps regulated by the model proposed in the present study coupling CTCF-mediated promoter-exon upstream looping and CTCF-mediated RNA Pol II pausing at differentially used exons. Further, we validated the expression and alternative splicing event of one of the model genes, *AOPEP*, in control and CTCF-depleted cells under hypoxia to strengthen our model mechanism. In corroboration with *COL5A1*, the *AOPEP* promoter showed CTCF enrichment in hypoxic cells compared to the normoxic cells (Figure 7F). As observed in CTCF-ChIP-seq, we further validated the CTCF oc-

cupancy upstream of alternatively spliced exon14 and found enriched CTCF occupancy at exon upstream locations (Figure 7F). We also observed enriched CTCF and RNA Pol II occupancy downstream of *AOPEP* exon14 under hypoxia in comparison to the normoxia control (Figure 7G). Additionally, we observed increased inclusion of exon14 under hypoxia in comparison to the normoxic control, which is reduced in CTCF-depleted hypoxic cells (Figure 7H). Additional illustrative examples of CTCF-related alternative splicing models are shown in Figures S7B–S7F. Overall, this analysis strengthens the model presented in this study, integrating the multiple layers of alternative splicing regulated by CTCF that promote EMT in breast cancer cells under hypoxic conditions.

DISCUSSION

Hypoxic niches present in the core of the solid tumor activate HIF pathways to initiate the EMT by reprogramming the epigenetics, chromatin structure of the genome, gene expression, and alternative splicing.^{3,28–33} We previously reported that hypoxia induces CTCF expression in an HIF1 α -dependent manner and that the HIF1 α -CTCF axis is one of the important axes that promotes the EMT.⁹

CTCF is a well-established master regulator of the genome, governing a wide array of genomic processes, including genome architecture, promoter-enhancer interactions, transcriptional regulation, long-range chromatin looping, and, more recently, alternative splicing, through its dynamic and widespread genomic occupancy.³⁴ In this study, we further investigated the specific process through which CTCF promotes the EMT under hypoxia. The data in the present study provide new insights into hypoxia-induced, CTCF-mediated regulation of gene expression and alternative splicing of metastatic genes. Our global analysis, utilizing CTCF CUT&RUN/ChIP-seq and RNA-seq data, suggests that CTCF potentially governs the expression of numerous genes and alternative splicing events associated with metastasis-related genes. Interestingly, we observed significant and widespread gain in CTCF occupancy in response to hypoxia, corroborating with the induced expression of CTCF under hypoxia.⁹ We demonstrated that hypoxia-induced CTCF enrichment at the *COL5A1* promoter leads to enhanced expression, which is facilitated by HIF1 α binding. Intriguingly, we found that HIF1 α occupancy on the *COL5A1* promoter is contingent upon CTCF binding, suggesting a potential connection between the two proteins, possibly controlling the transcription of other HIF1 α target genes.

(B) Schematic depicts genomic view of *COL5A1* gene loci on chromosome 9 and various putative convergent CTCF sites present in the promoter and proximal to exon64. The positions of the CTCF (yellow), HindIII sites (black arrow), and primers are indicated. C indicates the 3C anchor site, and T1, T2, and T3 indicate test sites.

(C) 3C analysis showing DNA-looping interactions between different HindIII fragments using an anchor primer C with test primer T1 and normalized 3C signal intensity for potential promoter-T1 interaction sites in CTCF-depleted MCF7 cells vs. shControl hypoxic cells in comparison to the normoxic control cells. Primers amplifying the genomic region with no HindIII cut site were used as an internal genomic control.

(D) CTCF-ChIP-qPCR at putative convergent CTCF motif (T1–T3) proximal to exon64 in normoxic and hypoxic MCF7 cells.

(E) Schematic representation of targeting the CTCF binding site involved in looping by dCas9-DNMT3A-sgRNA.

(F) MeDIP and CTCF ChIP-qPCR at T1 site in MCF7 cells transfected with dCas9-DNMT3A-sgCTCF_T1 vs. dCas9-DNMT3A-sgcontrol under hypoxia.

(G and H) 3C signal intensity using *COL5A1* promoter anchor primer and T1 site test primer (G) and MeDIP-qPCR, CTCF, and RNA Pol II ChIP-qPCR on *COL5A1* exon64A and RT-qPCR to check exon64A inclusion (H) in MCF7 cells transfected with dCas9-DNMT3A-CTCF_T1 vs. dCas9-DNMT3A-sgcontrol under hypoxia. Error bars, mean \pm SEM; two-tailed t test; * p < 0.05, ** p < 0.01, and *** p < 0.001; n = 3 biological replicates.

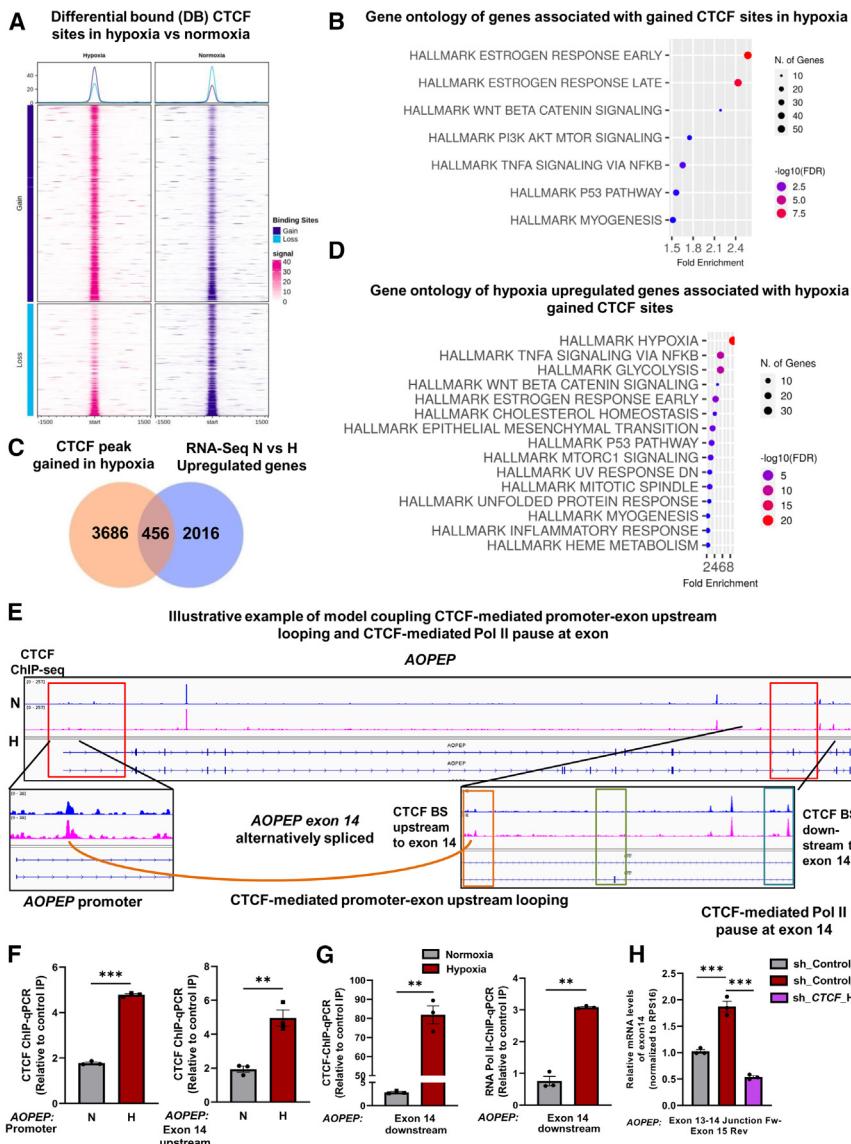


Figure 7. Hypoxia-induced differential CTCF occupancy regulates transcriptional and alternative splicing dynamics in breast cancer cells to drive EMT

(A) Differential binding analysis of CTCF occupancy inferred from CTCF-ChIP-seq between hypoxic and normoxic MCF7 cells. The top 1,000 differentially bound sites are shown.

(B) Representative significantly enriched terms from Gene Ontology and the curated MSigDB of genes associated with gained CTCF sites in hypoxia.

(C) Venn diagrams showing the overlap between hypoxia-gained CTCF sites (from A) and hypoxia-upregulated genes (from GEO: GSE166203).

(D) Representative significantly enriched terms from Gene Ontology and the curated MSigDB of hypoxia-upregulated genes associated with hypoxia-gained CTCF sites.

(E) AOPEP gene as an illustrative example of exon inclusion under hypoxia by model coupling CTCF-mediated promoter-exon upstream looping and CTCF-mediated RNA Pol II pausing at the exon shown in the present study.

(F) CTCF-ChIP-qPCR on AOPEP promoter and at putative convergent CTCF motif upstream of exon 14 in normoxic and hypoxic MCF7 cells.

(G) CTCF and RNA Pol II ChIP-qPCR on AOPEP exon14 downstream in normoxic and hypoxic MCF7 cells.

(H) RT-qPCR analysis showing AOPEP exon14 inclusion levels in MCF7 cells transduced with either shCTCF or shControl under hypoxia in comparison to the normoxia shControl.

Error bars, mean \pm SEM; two-tailed t test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$; $n = 3$ biological replicates.

rescued invasiveness in CTCF-depleted hypoxic breast cancer cells, while the COL5A1exon64B isoform did not. This finding highlights the critical role of the CTCF-COL5A1exon64A axis in the hypoxia-mediated invasive phenotype.

To date, numerous studies have highlighted the role of DNA methylation and CTCF-mediated RNA Pol II elongation-related mechanisms in the regulation of alternative splicing.^{19,35,36} In separate investigations, authors illustrated that CTCF-mediated RNA Pol II pausing at exons with weak splice sites promotes their inclusion, which is controlled by DNA methylation. More recently, our lab has demonstrated CTCF's involvement in regulating VEGFA splicing, where CTCF binding is significantly influenced by the presence or absence of the splicing factor SRSF2.³⁶ Notably, our study showed that CTCF occupancy on VEGFA-exon8a is independent of hypoxia and relies instead on SRSF2-dependent DNA methylation and that depletion of SRSF2 in normoxic conditions led to reduced DNA methylation and increased CTCF occupancy on VEGFA-exon8a.³⁶ Furthermore, the correlation between the intragenic CTCF binding sites, particularly those located proximal upstream of splice junctions, and alternative exon inclusion has been reported.²⁴

We further extended our investigation into how COL5A1 promotes the EMT. Interestingly, we found that hypoxia not only regulates COL5A1 expression but also modulates alternative splicing events in exon64. Hypoxic breast cancer cells prefer to include exon64A in COL5A1 mRNA. One of the outstanding questions in splicing biology is the mechanism underlying splicing specificity, which has implications for novel therapeutic strategies. Given CTCF's known role in mediating alternative splicing,^{19,34} we sought to determine its role in regulating the alternative splicing of COL5A1exon64A. Our study provides strong evidence that hypoxia-induced DNA demethylation and subsequent CTCF enrichment at exon64A mediate RNA Pol II pausing, favoring the inclusion of exon64A in COL5A1 mRNA and, thereby, promoting the EMT. Using the dCas9-DNMT3A-mediated targeted DNA methylation approach, we demonstrated a shift in alternative splicing, favoring exon64B inclusion over exon64A under hypoxia. Furthermore, restoration of the COL5A1exon64A isoform

Nonetheless, in what ways chromatin organization mediates alternative splicing is an important aspect yet not fully explored. Moreover, the interplay between the CTCF-mediated mechanisms that regulate alternative splicing, especially the impact of chromatin looping on CTCF-mediated RNA Pol II pausing and exon inclusion, remains largely uncharacterized. We found that hypoxia-driven CTCF enrichment occurs on the *COL5A1* promoter and, more surprisingly, upstream of the alternatively spliced exon64A, leading to promoter-alternate exon upstream looping. So, we expanded our study to decode the influence of CTCF-mediated promoter-alternate exon upstream looping on the outcome of alternative splicing. Our data revealed that hypoxia-driven, CTCF-mediated promoter-alternate exon upstream looping influences DNA demethylation on the distal exon64A. This regulation, in turn, governs hypoxia-mediated CTCF enrichment, RNA Pol II pausing at exon64A, and, consequently, CTCF-regulated exon64A inclusion.

In conclusion, we provide a model that shows how the hypoxia-CTCF axis coordinates the expression and alternative splicing of EMT genes, hence accelerating the EMT. We used epigenome editing to specifically impair CTCF binding, focusing on the HIF1 α -CTCF-COL5A1exon64A axis. This intervention reduces breast cancer cells' EMT capacity under hypoxia and may be a new therapeutic target in breast cancer treatment.

Limitations of the study

In this study, we identified significant and widespread gain in CTCF occupancy in response to hypoxia. This study sheds light on hypoxia-CTCF axis-mediated metastatic gene expression regulation and alternative splicing. Our study highlighted how the CTCF-mediated promoter-alternate exon upstream loop affects RNA Pol II pausing to mediate alternative splicing, specifically via regulating the DNA demethylation of distal exon64A of *COL5A1* and its role in the EMT under hypoxia. Nonetheless, the precise molecular mechanisms by which the CTCF-mediated promoter-alternate exon upstream loop influences the epigenetic reprogramming of downstream alternatively spliced exons remains an open question for future research. Further study is needed on how CTCF-mediated global chromatin looping regulates alternative splicing. While CTCF is recognized as a master regulator of chromatin organization, the impact of hypoxia-induced CTCF redistribution on genome architecture warrants further investigation.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Sanjeev Shukla (sanjeevs@iiserb.ac.in).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact without restriction.

Data and code availability

- CTCF-CUT&RUN, CTCF-ChIP-seq, and 4C-seq data have been deposited at GEO under GEO: GSE216843 and are publicly available as of the date of publication.

- Original western blot images have been deposited at Mendeley at <https://doi.org/10.17632/d2kvsgym55.2> and are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

P.K. is a recipient of a post-doctoral fellowship from the Indian Institute of Science Education and Research Bhopal, India. S.G.D. is supported by the Department of Biotechnology (DBT, India). D.P. is a recipient of funding from the University Grants Commission, India. A.P. is supported by the Prime Minister Research Fellowship (DST, India). This work is funded by a grant from the Department of Biotechnology (BT/PR44309/MED/30/2364/2021) and a Science and Engineering Research Board (SERB) grant (CRG/2021/004949).

AUTHOR CONTRIBUTIONS

Conceptualization, S.S. and P.K.; methodology, P.K. and S.S.; investigation, P.K., S.G.D., R.J., A.P., and S.S.; formal analysis, P.K., D.P., S.M., and D.N.; writing – original draft, P.K. and S.S.; funding acquisition, S.S.; resources, P.K. and S.S.; visualization, P.K., D.P., and S.S.; supervision, S.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Cell culture and treatment
- **METHOD DETAILS**
 - Generation of *COL5A1* promoter construct
 - Site-directed mutagenesis
 - Luciferase reporter assays
 - Molecular cloning of full length over expression constructs
 - Genomic DNA isolation and methylated DNA immunoprecipitation (MeDIP-qPCR)
 - Chromatin immunoprecipitation
 - Chromosome conformation capture (3C) assay
 - Transwell invasion assays
 - 3D spheroid invasion assay
 - Wound healing and cell counting Kit-8 (CCK-8) viability assays
 - Western blot
 - RNA isolation, cDNA synthesis and quantitative reverse transcription real-time PCR (qRT-PCR)
 - RNA interference
 - sgRNA designing and cloning in dCas9-DNMT3A expression vector and cell transfections
 - CRISPR-Cas9 mediated disruption of CTCF binding motifs
 - GSEA analysis
 - *COL5A1* expression analysis in publicly available dataset
 - *COL5A1* promoter region's in silico analysis for transcription factor binding sites and CpG islands
 - Retrieval of gene list from various databases
 - In silico analysis of CTCF binding motifs for the *COL5A1* locus
 - RNA-seq data analysis
 - ChIP-sequencing
 - ChIP-seq data analyses
 - CUT&RUN assay
 - Circular chromosome conformation capture (4C)
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2025.115267>.

Received: July 27, 2024

Revised: December 16, 2024

Accepted: January 15, 2025

Published: February 4, 2025

REFERENCES

1. Chaffer, C.L., and Weinberg, R.A. (2011). A perspective on cancer cell metastasis. *Science* 331, 1559–1564. <https://doi.org/10.1126/science.1203543>.
2. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 71, 209–249. <https://doi.org/10.3322/caac.21660>.
3. Hapke, R.Y., and Haake, S.M. (2020). Hypoxia-induced epithelial to mesenchymal transition in cancer. *Cancer Lett.* 487, 10–20. <https://doi.org/10.1016/j.canlet.2020.05.012>.
4. Muñoz, B., de la Puent, P., Azab, F., and Azab, A.K. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia* 3, 83–92. <https://doi.org/10.2147/hp.s93413>.
5. Rankin, E.B., and Giaccia, A.J. (2016). Hypoxic control of metastasis. *Science* 352, 175–180. <https://doi.org/10.1126/science.aaf4405>.
6. Semenza, G.L. (2012). Molecular mechanisms mediating metastasis of hypoxic breast cancer cells. *Trends Mol. Med.* 18, 534–543. <https://doi.org/10.1016/j.molmed.2012.08.001>.
7. LaGory, E.L., and Giaccia, A.J. (2016). The ever-expanding role of HIF in tumour and stromal biology. *Nat. Cell Biol.* 18, 356–365. <https://doi.org/10.1038/ncb3330>.
8. Yang, M.H., Wu, M.Z., Chiou, S.H., Chen, P.M., Chang, S.Y., Liu, C.J., Teng, S.C., and Wu, K.J. (2008). Direct regulation of TWIST by HIF-1 α promotes metastasis. *Nat. Cell Biol.* 10, 295–305. <https://doi.org/10.1038/ncb1691>.
9. Kakani, P., Dhamdhare, S.G., Pant, D., Joshi, R., Mishra, J., Samaiya, A., and Shukla, S. (2024). Hypoxia-induced CTCF promotes EMT in breast cancer. *Cell Rep.* 43, 114367. <https://doi.org/10.1016/j.celrep.2024.114367>.
10. Buyuk, B., Jin, S., and Ye, K. (2022). Epithelial-to-Mesenchymal Transition Signaling Pathways Responsible for Breast Cancer Metastasis. *Cell. Mol. Bioeng.* 15, 1–13. <https://doi.org/10.1007/s12195-021-00694-9>.
11. Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* 10, 1038.
12. Gröger, C.J., Grubinger, M., Waldhör, T., Vierlinger, K., and Mikulits, W. (2012). Meta-Analysis of Gene Expression Signatures Defining the Epithelial to Mesenchymal Transition during Cancer Progression. *PLoS One* 7, e51136. <https://doi.org/10.1371/journal.pone.0051136>.
13. Foroutan, M., Cursors, J., Hediyyeh-Zadeh, S., Thompson, E.W., and Davis, M.J. (2017). A transcriptional program for detecting TGF β -induced EMT in cancer. *Mol. Cancer Res.* 15, 619–631. <https://doi.org/10.1158/1541-7786.MCR-16-0313>.
14. Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T., Zhang, W., Yu, X., and Liu, L. (2019). The role of collagen in cancer: From bench to bedside. *J. Transl. Med.* 17, 309. <https://doi.org/10.1186/s12967-019-2058-1>.
15. Zhang, Z., Wang, Y., Zhang, J., Zhong, J., and Yang, R. (2018). COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol. Med. Rep.* 17, 5037–5042. <https://doi.org/10.3892/mmr.2018.8533>.
16. Gu, J., Lu, Y., Li, F., Qiao, L., Wang, Q., Li, N., Borgia, J.A., Deng, Y., Lei, G., and Zheng, Q. (2014). Identification and characterization of the novel Col10a1 regulatory mechanism during chondrocyte hypertrophic differentiation. *Cell Death Dis.* 5, e1469. <https://doi.org/10.1038/cddis.2014.444>.
17. Li, T., Huang, H., Shi, G., Zhao, L., Li, T., Zhang, Z., Liu, R., Hu, Y., Liu, H., Yu, J., and Li, G. (2018). TGF- β 1-SOX9 axis-inducible COL10A1 promotes invasion and metastasis in gastric cancer via epithelial-to-mesenchymal transition. *Cell Death Dis.* 9, 849. <https://doi.org/10.1038/s41419-018-0877-2>.
18. Marina, R.J., Sturgill, D., Bailly, M.A., Thenoz, M., Varma, G., Prigge, M.F., Nanan, K.K., Shukla, S., Haque, N., and Oberdoerffer, S. (2016). TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.* 35, 335–355. <https://doi.org/10.1525/embj.201593235>.
19. Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79. <https://doi.org/10.1038/nature10442>.
20. D'Anna, F., Van Dyck, L., Xiong, J., Zhao, H., Berrens, R.V., Qian, J., Bieniasz-Krzywiec, P., Chandra, V., Schoonjans, L., Matthews, J., et al. (2020). DNA methylation repels binding of hypoxia-inducible transcription factors to maintain tumor immunotolerance. *Genome Biol.* 21, 182. <https://doi.org/10.1186/s13059-020-02087-z>.
21. Mitchell, A.L., Judis, L.M., Schwarze, U., Vaynshtok, P.M., Drumm, M.L., and Byers, P.H. (2012). Characterization of tissue-specific and developmentally regulated alternative splicing of exon 64 in the COL5A1 gene. *Connect. Tissue Res.* 53, 267–276. <https://doi.org/10.3109/03008207.2011.636160>.
22. Hoffman, G.G., Branam, A.M., Huang, G., Pelegri, F., Cole, W.G., Wenstrup, R.M., and Greenspan, D.S. (2010). Characterization of the six zebrafish clade B fibrillar procollagen genes, with evidence for evolutionarily conserved alternative splicing within the pro- α 1(V) C-propeptide. *Matrix Biol.* 29, 261–275. <https://doi.org/10.1016/j.matbio.2010.01.006>.
23. López Soto, E.J., and Lipscombe, D. (2020). Cell-specific exon methylation and CTCF binding in neurons regulate calcium ion channel splicing and function. *Elife* 9, 54879. <https://doi.org/10.7554/elife.2023.54879>.
24. Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.M., Gibson, T.J., and Zaugg, J.B. (2017). CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. *Cell Syst.* 5, 628–637.e6. <https://doi.org/10.1016/j.cels.2017.10.018>.
25. Acuña, L.I.G., Flyamer, I., Boyle, S., Friman, E., and Bickmore, W.A. (2023). Transcription decouples estrogen-dependent changes in enhancer-promoter contact frequencies and spatial proximity. *Preprint at bioRxiv 2023*, 534720. <https://doi.org/10.1101/2023.03.29.534720>.
26. de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell* 60, 676–684. <https://doi.org/10.1016/j.molcel.2015.09.023>.
27. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900–910. <https://doi.org/10.1016/j.cell.2015.07.038>.
28. Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., Ghesquière, B., Van Dyck, L., Boeckx, B., Schoonjans, L., et al. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* 537, 63–68. <https://doi.org/10.1038/nature19081>.
29. Rankin, E.B., Nam, J.M., and Giaccia, A.J. (2016). Hypoxia: Signaling the Metastatic Cascade. *Trends Cancer* 2, 295–304. <https://doi.org/10.1016/j.trecan.2016.05.006>.
30. Farina, A.R., Cappabianca, L., Sebastian, M., Zelli, V., Guadagni, S., and Mackay, A.R. (2020). Hypoxia-induced alternative splicing: The 11th Hallmark of Cancer. *J. Exp. Clin. Cancer Res.* 39, 110. <https://doi.org/10.1186/s13046-020-01616-9>.

31. Tsai, Y.P., and Wu, K.J. (2014). Epigenetic regulation of hypoxia-responsive gene expression: Focusing on chromatin and DNA modifications. *Int. J. Cancer* 134, 249–256. <https://doi.org/10.1002/ijc.28190>.
32. Prickaerts, P., Adriaens, M.E., Beucken, T.v.d., Koch, E., Dubois, L., Dahlmans, V.E.H., Gits, C., Evelo, C.T.A., Chan-Seng-Yue, M., Wouters, B.G., and Voncken, J.W. (2016). Hypoxia increases genome-wide bivalent epigenetic marking by specific gain of H3K27me3. *Epigenet. Chromatin* 9, 46. <https://doi.org/10.1186/s13072-016-0086-0>.
33. Ye, L.Y., Chen, W., Bai, X.L., Xu, X.Y., Zhang, Q., Xia, X.F., Sun, X., Li, G.G., Hu, Q.D., Fu, Q.H., and Liang, T.B. (2016). Hypoxia-induced epithelial-to-mesenchymal transition in hepatocellular carcinoma induces an immunosuppressive tumor microenvironment to promote metastasis. *Cancer Res.* 76, 818–830. <https://doi.org/10.1158/0008-5472.CAN-15-0977>.
34. Alharbi, A.B., Schmitz, U., Bailey, C.G., and Rasko, J.E.J. (2021). CTCF as a regulator of alternative splicing: New tricks for an old player. *Nucleic Acids Res.* 49, 7825–7838. <https://doi.org/10.1093/nar/gkab520>.
35. López Soto, E.J., and Lipscombe, D. (2020). Cell-specific exon methylation and CTCF binding in neurons regulate calcium ion channel splicing and function. *eLife* 9, 54879. <https://doi.org/10.7554/eLife.2023.106804>.
36. Yadav, P., Pandey, A., Kakani, P., Mutnuru, S.A., Samaiya, A., Mishra, J., and Shukla, S. (2023). Hypoxia-induced loss of SRSF2-dependent DNA methylation promotes CTCF-mediated alternative splicing of VEGFA in breast cancer. *iScience* 26, 106804. <https://doi.org/10.1016/j.isci.2023.106804>.
37. Zhang, X., Luo, F., Luo, S., Li, L., Ren, X., Lin, J., Liang, Y., Ma, C., Ding, L., Zhang, D., et al. (2022). Transcriptional Repression of Aerobic Glycolysis by OVO1 in Breast Cancer. *Adv. Sci.* 9, e2200705. <https://doi.org/10.1002/advs.202200705>.
38. Ahuja, N., Ashok, C., Natua, S., Pant, D., Cherian, A., Pandkar, M.R., Yadav, P., Vishnu, N.S.S., Mishra, J., Samaiya, A., and Shukla, S. (2020). Hypoxia-induced TGF-β-RBFOX2-ESRP1 axis regulates human MENA alternative splicing and promotes EMT in breast cancer. *NAR Cancer* 2, zcaa021. <https://doi.org/10.1093/narcan/zcaa021>.
39. Liu, Y.R., Jiang, Y.Z., Xu, X.E., Hu, X., Yu, K.D., and Shao, Z.M. (2016). Comprehensive transcriptome profiling reveals multigene signatures in triple-negative breast cancer. *Clin. Cancer Res.* 22, 1653–1662. <https://doi.org/10.1158/1078-0432.CCR-15-1555>.
40. Vojta, A., Dobrinić, P., Tadic, V., Bockor, L., Korac, P., Julg, B., Klasic, M., and Zoldos, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* 44, 5615–5628. <https://doi.org/10.1093/nar/gkw159>.
41. Sikder, R.K., Ellithi, M., Uzzo, R.N., Weader, D.J., Metz, A.L., Behbahani, A., McKenzie, E.R., El-Deiry, W.S., and Abbosh, P.H. (2021). Differential effects of clinically relevant N- versus C-terminal truncating CDKN1A mutations on cisplatin sensitivity in bladder cancer. *Mol. Cancer Res.* 10, 1158/1541–7786. MCR-19-1200.
42. Dreos, R., Ambrosini, G., Périer, R.C., and Bucher, P. (2015). The Eukaryotic Promoter Database: Expansion of EPDNew and new promoter analysis tools. *Nucleic Acids Res.* 43, D92–D96. <https://doi.org/10.1093/nar/gku1111>.
43. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
44. Meylan, P., Dreos, R., Ambrosini, G., Groux, R., and Bucher, P. (2020). EPD in 2020: Enhanced data visualization and extension to ncRNA promoters. *Nucleic Acids Res.* 48, D65–D69. <https://doi.org/10.1093/nar/gkz1014>.
45. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. <https://doi.org/10.1038/nbt.1754>.
46. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
47. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
48. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389–393. <https://doi.org/10.1038/nature10730>.
49. Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kasprzyk, A. (2009). BioMart - Biological queries made easy. *BMC Genom.* 10, 22. <https://doi.org/10.1186/1471-2164-10-22>.
50. Ge, S.X., Jung, D., and ShinyGO, R.Y. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>.
51. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
52. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* 111, E5593–E5601. <https://doi.org/10.1073/pnas.1419161111>.
53. Tarazona, S., García, F., Ferrer, A., Dopazo, J., and Conesa, A. (2012). NOIseq: a RNA-seq differential expression method robust for sequencing depth biases. *EMBnet. journal* 17, 18–19. <https://doi.org/10.14806/ej.17.b.265>.
54. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>.
55. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110–D115. <https://doi.org/10.1093/nar/gkv1176>.
56. Fornes, O., Castro-Mondragon, J.A., Khan, A., Van Der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranášić, D., et al. (2020). JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. *Nucleic Acids Res.* 48, D87–D92. <https://doi.org/10.1093/nar/gkz1001>.
57. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
58. Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 47, W556–W560. <https://doi.org/10.1093/nar/gkz430>.
59. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. <https://doi.org/10.1016/j.cell.2014.08.009>.
60. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16, 276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2).
61. Kuo, H.C., Lin, P.Y., Chung, T.C., Chao, C.M., Lai, L.C., Tsai, M.H., and Chuang, E.Y. (2011). DBCAT: Database of CpG islands and analytical tools for identifying comprehensive methylation profiles in cancer cells. *J. Comput. Biol.* 18, 1013–1017. <https://doi.org/10.1089/cmb.2010.0038>.

62. Conant, D., Hsiau, T., Rossi, N., Oki, J., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., et al. (2022). Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J.* 5, 123–130. <https://doi.org/10.1089/crispr.2021.0113>.
63. Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475. <https://doi.org/10.1093/bioinformatics/btu048>.
64. O'Brien, A., and Bailey, T.L. (2014). GT-Scan: Identifying unique genomic targets. *Bioinformatics* 30, 2673–2675. <https://doi.org/10.1093/bioinformatics/btu354>.
65. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vainberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. <https://doi.org/10.1038/nbt.3437>.
66. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
67. Van De Werken, H.J.G., Landan, G., Holwerda, S.J.B., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Öz, Y., Bouwman, B.A.M., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* 9, 969–972. <https://doi.org/10.1038/nmeth.2173>.
68. Krijger, P.H.L., Geeven, G., Bianchi, V., Hilvering, C.R.E., and de Laat, W. (2020). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* 170, 17–32. <https://doi.org/10.1016/jymeth.2019.07.014>.
69. Pandey, A., Kakani, P., and Shukla, S. (2024). CTCF and BORIS-mediated autophagy regulation via alternative splicing of BNIP3L in breast cancer. *J. Biol. Chem.* 300, 107416. <https://doi.org/10.1016/j.jbc.2024.107416>.
70. Singh, S., Narayanan, S.P., Biswas, K., Gupta, A., Ahuja, N., Yadav, S., Panday, R.K., Samaiya, A., Sharan, S.K., and Shukla, S. (2017). Intragenic DNA methylation and BORIS-mediated cancer-specific splicing contribute to the Warburg effect. *Proc. Natl. Acad. Sci. USA* 114, 11440–11445. <https://doi.org/10.1073/pnas.1708447114>.
71. Ashok, C., Ahuja, N., Natua, S., Mishra, J., Samaiya, A., and Shukla, S. (2021). E2F1 and epigenetic modifiers orchestrate breast cancer progression by regulating oxygen-dependent ESRP1 expression. *Oncogenesis* 10, 58. <https://doi.org/10.1038/s41389-021-00347-6>.
72. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods* 25, 402–408. <https://doi.org/10.1006/meth.2001.1262>.
73. Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., De Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3c-qpcr). *Nat. Protoc.* 2, 1722–1733. <https://doi.org/10.1038/nprot.2007.243>.
74. Van De Werken, H.J.G., De Vree, P.J.P., Splinter, E., Holwerda, S.J.B., Klous, P., De Wit, E., and De Laat, W. (2012). 4C technology: Protocols and data analysis. *Methods Enzymol.* 513, 89–112. <https://doi.org/10.1016/B978-0-12-391938-0-00004-5>.
75. Gupta, A., Yadav, S., PT, A., Mishra, J., Samaiya, A., Panday, R.K., and Shukla, S. (2020). The HNRNPA2B1-MST1R-Akt axis contributes to epithelial-to-mesenchymal transition in head and neck cancer. *Lab. Invest.* 100, 1589–1601. <https://doi.org/10.1038/s41374-020-0466-8>.
76. Li, S.Y., Hammarlund, J.A., Wu, G., Lian, J.W., Howell, S.J., Clarke, R.B., Adamson, A.D., Gonçalves, C.F., Hogenesch, J.B., Anafi, R.C., and Meng, Q.J. (2024). Tumor circadian clock strength influences metastatic potential and predicts patient prognosis in luminal A breast cancer. *Proc. Natl. Acad. Sci. USA* 121, e2311854121. <https://doi.org/10.1073/pnas.2311854121>.
77. Padmanaban, V., Krol, I., Suhail, Y., Szczzerba, B.M., Aceto, N., Bader, J.S., and Ewald, A.J. (2019). E-cadherin is required for metastasis in multiple models of breast cancer. *Nature* 573, 439–444. <https://doi.org/10.1038/s41586-019-1526-3>.
78. Grada, A., Otero-Vinas, M., Prieto-Castrillo, F., Obagi, Z., and Falanga, V. (2017). Research Techniques Made Simple: Analysis of Collective Cell Migration Using the Wound Healing Assay. *J. Invest. Dermatol.* 137, e11–e16. <https://doi.org/10.1016/j.jid.2016.11.020>.
79. Gupta, A., Ajith, A., Singh, S., Panday, R.K., Samaiya, A., and Shukla, S. (2018). PAK2-c-Myc-PKM2 axis plays an essential role in head and neck oncogenesis via regulating Warburg effect. *Cell Death Dis.* 9, 825. <https://doi.org/10.1038/s41419-018-0887-0>.
80. Yadav, S., Pant, D., Samaiya, A., Kalra, N., Gupta, S., and Shukla, S. (2021). ERK1/2-EGFR-SRSF10 Axis Mediated Alternative Splicing Plays a Critical Role in Head and Neck Cancer. *Front. Cell Dev. Biol.* 9, 713661. <https://doi.org/10.3389/fcell.2021.713661>.
81. Tang, Q., Chen, J., Di, Z., Yuan, W., Zhou, Z., Liu, Z., Han, S., Liu, Y., Ying, G., Shu, X., and Di, M. (2020). TM4SF1 promotes EMT and cancer stemness via the Wnt/β-catenin/SOX2 pathway in colorectal cancer. *J. Exp. Clin. Cancer Res.* 39, 232. <https://doi.org/10.1186/s13046-020-01690-z>.
82. Suarez-Arnedo, A., Figueroa, F.T., Clavijo, C., Arbeláez, P., Cruz, J.C., and Muñoz-Camargo, C. (2020). An image J plugin for the high throughput image analysis of in vitro scratch wound healing assays. *PLoS One* 15, 0232565. <https://doi.org/10.1371/journal.pone.0232565>.
83. Yue, P.Y.K., Leung, E.P.Y., Mak, N.K., and Wong, R.N.S. (2010). A Simplified Method for Quantifying Cell Migration/Wound Healing in 96-Well Plates. *J. Biomol. Screen* 15, 427–433. <https://doi.org/10.1177/1087077110361772>.
84. Pandkar, M.R., Raveendran, A., Biswas, K., Mutnuru, S.A., Mishra, J., Samaiya, A., Malys, T., Mitrophanov, A.Y., Sharan, S.K., and Shukla, S. (2023). PKM2 dictates the poised chromatin state of PFKFB3 promoter to enhance breast cancer progression. *NAR Cancer* 5, zcad032. <https://doi.org/10.1093/narcan/zcad032>.
85. Pandkar, M.R., Sinha, S., Samaiya, A., and Shukla, S. (2023). Oncometabolite lactate enhances breast cancer progression by orchestrating histone lactylation-dependent c-Myc expression. *Transl. Oncol.* 37, 101758. <https://doi.org/10.1016/j.tranon.2023.101758>.
86. Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* 2, 16009. <https://doi.org/10.1038/celdisc.2016.9>.
87. Pant, D., Narayanan, S.P., Vijay, N., and Shukla, S. (2020). Hypoxia-induced changes in intragenic DNA methylation correlate with alternative splicing in breast cancer. *J. Bio. Sci.* 45, 3. <https://doi.org/10.1007/s12038-019-9977-0>.
88. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* 1, 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>.
89. Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profilier, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14, 482–517. <https://doi.org/10.1038/s41596-018-0103-9>.
90. Sarrió, D., Rodriguez-Pinilla, S.M., Hardisson, D., Cano, A., Moreno-Bueno, G., and Palacios, J. (2008). Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res.* 68, 989–997. <https://doi.org/10.1158/0008-5472.CAN-07-2017>.
91. Jechlinger, M., Grunert, S., Tamir, I.H., Janda, E., Lüdemann, S., Waerner, T., Seither, P., Weith, A., Beug, H., and Kraut, N. (2003). Expression profiling of epithelial plasticity in tumor progression. *Oncogene* 22, 7155–7169. <https://doi.org/10.1038/sj.onc.1206887>.

92. Cheng, Q., Chang, J.T., Gwin, W.R., Zhu, J., Ambs, S., Geradts, J., and Lyerly, H.K. (2014). A signature of epithelial-mesenchymal plasticity and stromal activation in primary tumor modulates late recurrence in breast cancer independent of disease subtype. *Breast Cancer Res.* 16, 407. <https://doi.org/10.1186/s13058-014-0407-9>.
93. Taube, J.H., Herschkowitz, J.I., Komurov, K., Zhou, A.Y., Gupta, S., Yang, J., Hartwell, K., Onder, T.T., Gupta, P.B., Evans, K.W., et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. USA* 107, 15449–15454. <https://doi.org/10.1073/pnas.1004900107>.
94. Sharma, S., Chung, C.Y., Uryu, S., Petrovic, J., Cao, J., Rickard, A., Nady, N., Greasley, S., Johnson, E., Brodsky, O., et al. (2023). Discovery of a highly potent, selective, orally bioavailable inhibitor of KAT6A/B histone acetyltransferases with efficacy against KAT6A-high ER+ breast cancer. *Cell Chem. Biol.* 30, 1191–1210.e20. <https://doi.org/10.1016/j.chembiol.2023.07.005>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CTCF (D31H2)	Cell Signaling Technology	Cat# 3418S; Lot# 1; RRID:AB_2086791
HIF-1 α (D2U3T)	Cell Signaling Technology	Cat# 14179S; Lot# 3; RRID:AB_2622225
GAPDH (D16h11)	Cell Signaling Technology	Cat# 5174S; Lot# 8; RRID:AB_10622025
COL5A1	Cell Signaling Technology	Cat# 37304S; Lot# 1; RRID:AB_3675529
Anti-E cadherin	Abcam	Cat# ab40772; Lot# GR148899-1; RRID:AB_731493
Anti-Vimentin	Abcam	Cat# ab137321; Lot# GR294886-5; RRID:AB_2921312
Alexa-Flour 680 anti-rabbit IgG	Invitrogen	Cat# A32734; Lot# RJ243414; RRID:AB_2633283
Alexa-Flour 800 anti-mouse IgG	Invitrogen	Cat# A32730; Lot# SC243837; RRID:AB_2633279
Cas9 (7A9-3A3)	Cell Signaling Technology	Cat# 14697S; Lot# 3, RRID:AB_2750916
Rpb1 CTD (4H8) (RNA Polymerase II CTD)	Cell Signaling Technology	Cat# 2629S; Lot# 3; RRID:AB_2167468
5-methylCytosine (D3s2z)	Cell Signaling Technology	Cat# 28692S; Lot# 2; RRID:AB_2798962
Normal Rabbit IgG	Cell Signaling Technology	Cat# 2729S; Lot# 10; RRID:AB_1031062
Normal Mouse IgG	Millipore	Cat# NI03-100UG; Lot# 3129279; RRID:AB_3675530
GFP-Tag Rabbit Polyclonal antibody	Affinity Bioscience	Cat# T0006; Lot# 56 × 8617; RRID:AB_2839423
Bacterial and virus strains		
<i>E. Coli</i> stbl3 cells	Thermo Fisher Scientific	Cat# C737303
Chemicals, peptides, and recombinant proteins		
Lipofectamine 2000	ThermoFisher Scientific	Cat# 11668019
TurboFect	Thermo scientific	Cat# R0531
Opti-MEM	Invitrogen	Cat# 31985062
Polybrene	Sigma	Cat# H9268
Puromycin	Sigma	Cat# P9620
Fetal Bovine Serum	Sigma	Cat# F7524
Penicillin and Streptomycin	Invitrogen	Cat# 15140122
L-Glutamine	Invitrogen	Cat# 25030081
TRIzol	Invitrogen	Cat# 15596018
Protein G Dynabeads	Invitrogen	Cat# 10004D
SYBR Green Master Mix	Promega	Cat# A6101
PrimeScript cDNA synthesis kit	TaKaRa	Cat# 6110A
Dream Taq DNA Pol	Thermo Fisher Scientific	EP0702
Phusion HF DNA Polymerase	New England Biolabs	Cat# M0530S
Proteinase K Solution (20 mg/mL), RNA grade	Invitrogen	Cat# 25530049
PureLink™ RNase A (20 mg/mL)	Thermo Fischer Scientific	Cat# 12091021
HindIII-HF	New England Biolabs	Cat# R3104T; Lot# 10029681
T4 DNA ligase	New England Biolabs	Cat# M0202S; Lot# 10113858
KpnI	TaKaRa	Cat# 1068A; Lot# AL50748A
D-Luciferin sodium salt	Cayman Chemical	Cat# 14682-50mg; Lot# 0639766-1
Coelenterazine	Cayman Chemical	Cat# 16123; Lot # 0560653-5
Matrigel	Corning	Cat# 356230; Lot # 2010001

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Expand Long Template Polymerase mix	Roche	Cat #11681834001
Corning® Collagen I, Rat Tail	Corning®	Cat #354236
DpnII	New England Biolabs	Cat #R0543T
NlaIII	New England Biolabs	Cat #R0125S
Spermidine	Sigma	Cat #S2626-1G
Digitonin	Sigma	Cat #D141-100MG
Concavalin-A	Cell Signaling Technology	Cat #93569S
pAG-MNase	Cell Signaling Technology	Cat #40366S
GlycoBlue™ Coprecipitant (15 mg/mL)	Thermo Fisher Scientific	AM9515
Critical commercial assays		
QIAprep Spin Miniprep Kit	Qiagen	Cat# 27106
QIAquick PCR purification Kit	Qiagen	Cat# 28106
GenElute Mammalian Genomic DNA Miniprep Kits	Sigma	Cat# G1N70
QIAquick Gel extraction Kit	Qiagen	Cat# 28706
DNA Ligation kit ver2.1	TaKaRa	Cat# 6022
NEBNext Ultra II DNA library Prep Kit for Illumina	New England Biolabs	Cat #E7645S
CELL COUNTING KIT - 8	Sigma	Cat #96992-500TESTS-F
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) – 24 reactions	New England Biolabs	Cat #E7335S
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 2)	New England Biolabs	Cat #E7500S
Deposited data		
CTCF-CUT&RUN assay in normoxic and hypoxic MCF7 cell line	This paper	GSE216843
CTCF-ChIP-seq in normoxic and hypoxic MCF7 cell line	This paper	GSE216843
RNA-seq in normoxic and hypoxic MCF7 cell line	Zhang et al. ³⁷	GSE166203
4C-seq	This paper	GSE216843
HTA in normoxic and hypoxic HCC1806 cell line	Ahuja et al. ³⁸	GSE147516
HTA in triple negative breast cancer patient sample	Liu et al. ³⁹	GSE76250
Unprocessed blots	This paper	Mendeley Data: https://doi.org/10.17632/d2kvsgym55.2
Experimental models: Cell lines		
MCF7	ATCC	Cat# HTB -22
HCC1806	ATCC	Cat# CRL-2335
HEK293T	ATCC	Cat# CRL-3216
Oligonucleotides		
Molecular Cloning Primers, See Table S4	This paper	N/A
ChIP-qPCR and MeDIP-qPCR Primers, See Table S5	This paper	N/A
3C and 4C Primers, See Table S5	This paper	N/A
qPCR Primers, See Table S6	This paper	N/A
shRNA, See Table S7	Sigma, Mission Human Genome shRNA Library	N/A
sgRNA, See Table S7	This paper	N/A
Off target primers, See Table S7	This paper	N/A
Recombinant DNA		
pGL3-Basic	Promega	E1751
pGL3-Control	Promega	E1741
pRL-TK Renilla luciferase plasmid	Promega	E2231

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pdCas9-DNMT3A-PuroR_v2	Vojta et al. ⁴⁰	Addgene #74407
pdCas9-DNMT3A-PuroR_v2-sgcontrol	Vojta et al. ⁴⁰	Addgene #71830
pLentiCRISPR-E	Sikder et al. ⁴¹	Addgene #78852
pEGFP-C3	Clontech	6082-1
pGL3-COL5A1-promoter(-337-+115)-WT	This paper	N/A
pGL3-COL5A1-promoter(-337-+115)-HRE-Mutant	This paper	N/A
pGL3-COL5A1-promoter(-337-+115)-CTCF-Mutant	This paper	N/A
pGL3-COL5A1-promoter(-337-+115)-HRE/CTCF-Mutant	This paper	N/A
pEGFP-C3-COL5A1-exon64A	This paper	N/A
pEGFP-C3-COL5A1-exon64B	This paper	N/A
pLentiCRISPR-E– COL5A1-CTCF-sgRNA1	This paper	N/A
pLentiCRISPR-E– COL5A1-CTCF-sgRNA2	This paper	N/A
pLentiCRISPR-E– COL5A1 Exon 64A-CTCF-SgRNA1	This paper	N/A
pLentiCRISPR-E– COL5A1 Exon 64A-CTCF-SgRNA2	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1 Exon 64A-SgRNA1	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1 Exon 64A-SgRNA2	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1 Exon 64A-SgRNA3	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1 Exon 64A-SgRNA4	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1_T1-sgRNA1	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1_T1-sgRNA2	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1_HRE_sgRNA	This paper	N/A
pdCas9-DNMT3A-PuroR_v2-COL5A1_CTCF_sgRNA	This paper	N/A

Software and algorithms

Eukaryotic Promoter Database	Dreos et al. ⁴²	https://epd.vital-it.ch/
MACS2 v2.1.2	Zhang et al. ⁴³	https://github.com/macs3-project/MACS/
Eukaryotic Promoter Database New	Meylan et al. ⁴⁴	https://epd.expasy.org/epd/EPDnew_database.php
IGV	Robinson et al. ⁴⁵	http://www.broadinstitute.org/igv/
Trimmomatic (v0.39)	Bolger et al. ⁴⁶	http://www.usadellab.org/cms/?page=trimmomatic
Samtools v1.9	Li et al. ⁴⁷	http://htslib.org/
DiffBind v3.6.5	Ross-Innes et al. ⁴⁸	http://bioconductor.org/packages/release/bioc/html/DiffBind.html
BioMart by ENSEMBL	Smedley et al. ⁴⁹	http://www.ensembl.org/info/data/biomart/biomart_r_package.html
ShinyGO (v 0.80)	Ge et al. ⁵⁰	http://bioinformatics.sdbstate.edu/go/
BEDtools (v2.26.0)	Quinlan and Hall, 2010 ⁵¹	https://bedtools.readthedocs.io/en/latest/
rMATS (v. 4.0.2)	Shen et al. ⁵²	https://rnaseq-mats.sourceforge.io/
NOIseq	Tarazona et al. ⁵³	https://www.bioconductor.org/packages/release/bioc/html/NOISeq.html
FIMO	Grant et al. ⁵⁴	https://meme-suite.org/meme/tools/fimo
JASPAR 2016 server	Mathelier et al. ⁵⁵	http://jaspar2016.genereg.net/
JASPAR 2020 server	Fornes et al. ⁵⁶	https://jaspar2020.genereg.net/
Bowtie 2, version 2.3.4	Langmead and Salzberg ⁵⁷	https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
GEPIA2	Tang et al. ⁵⁸	http://gepia2.cancer-pku.cn
CIS-BP	Weirauch et al. ⁵⁹	http://cisbp.ccbr.utoronto.ca/
EMBOSS CpGPlot	Rice et al. ⁶⁰	https://www.ebi.ac.uk/emboss/cpgplot

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DBCAT	Kuo et al. ⁶¹	http://dbcat.cgm.ntu.edu.tw
ICE analysis	Conant et al. ⁶²	https://ice.synthego.com
Cas-OFFinder	Bae et al. ⁶³	http://www.rgenome.net/cas-offinder
GT-Scan web-tool	O'Brien and Bailey ⁶⁴	http://gt-scan.braembl.org.au/gt-scan
CRISPRick	Doench et al. ⁶⁵	https://portals.broadinstitute.org/gppx/crispick/public
ImageJ software	NIH	https://imagej.nih.gov/ij/
Gene Set Enrichment Analysis (GSEA)	Subramanian et al. ⁶⁶	http://software.broadinstitute.org/gsea/
4Cseqpipe	van de Werken et al. ⁶⁷	http://compgenomics.weizmann.ac.il/tanay/?page_id=367
pipe4C	Krijger et al. ⁶⁸	https://github.com/deLaatLab/pipe4C
PRISM 8	GraphPad	https://www.graphpad.com/scientificsoftware/prism

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell culture and treatment

American Type Culture Collection (ATCC) provided the human breast cancer cell lines MCF7 and HCC1806, which were cultured in ATCC-recommended DMEM and RPMI1640, respectively. The media was supplemented with 10% fetal bovine serum (FBS; Sigma, F7524), 100 units/ml of penicillin and streptomycin (Invitrogen, 15140122), and 2 mmol/IL-glutamine (Sigma, G7513). Cells were cultivated in 5% CO₂ at 37°C. The Ruskinn INVIVO2 400 hypoxia chamber was used to keep the cells for treatment under hypoxic conditions (1% O₂) as before.^{9,38,69}

METHOD DETAILS

Generation of COL5A1 promoter construct

The human COL5A1 promoter sequence was obtained from the Eukaryotic Promoter Database (<https://epd.vital-it.ch/>⁴²) to generate the promoter construct. Using MCF7 genomic DNA as a template COL5A1 gene promoter using PCR was amplified from -337 bp upstream to +115 bp downstream of the TSS at +1, and then it was introduced into the pGL3-Basic expression vector (Promega, E1751). Primers are given in Table S4. The COL5A1 promoter construct was cloned between the KpnI F and HindIII R sites.

Site-directed mutagenesis

The wild-type -337 to +115 COL5A1 promoter luciferase construct was used as template to generate site-directed mutant of the COL5A1 promoter-luciferase construct using oligonucleotides (Table S4) with mutations in the HRE and CTCF binding site present from -227 to -219 (CCACGTCC to CCAAAACC) and -153 to -134 (CGCCCAGTGGGAGGCAGGG to CGTTCAAGAA AAGACAAAG) bp upstream of the COL5A1 TSS, respectively. The Mutations in the COL5A1 promoter delete construct was confirmed by DNA sequencing.

Luciferase reporter assays

Breast cancer cell lines MCF7 or HCC1806 (0.05×10^6 or 0.3×10^6) were seeded in 24-well plates, respectively and cultured for 16 h. The pRL-TK Renilla luciferase plasmid (Promega, E2231) and the COL5A1 promoter (-337 to +115) luciferase deletion construct were co-transfected into these cells and pGL3-control (Promega, E1741) was used as control plasmid. Post 12 h transfection, cells were kept in hypoxia chamber and passive lysis buffer (1% Triton X-100, 25mM Tricine pH 7.8, 15mM potassium phosphate pH 7.8, 15mM magnesium sulfate, 4mM EGTA, 1mM DTT) was used to lyse the cells. Using lysis buffer supplemented with Luciferin (Cayman Chemical, Cat no. 14682-50mg, Lot no. 0639766-1), the firefly luciferase activities were measured in a SpectraMax-Multi Detection System. The obtained luciferase values were normalized to Renilla luciferase activities (Cayman Chemical, Cat no. 16123, Lot no. 0560653-5).

Molecular cloning of full length over expression constructs

MCF7 cDNA was used as a template and COL5A1 (exon64A and exon64B isoform) CDS were amplified using phusion DNA polymerase (NEB, M0530S) and cloned in pEGFP-C3 (Addgene, 6082-1) overexpression plasmid. Primers are given in Table S4. COL5A1 exon64A and exon64B isoform was cloned between the HindIII F and KpnI R sites.

Genomic DNA isolation and methylated DNA immunoprecipitation (MeDIP-qPCR)

After treatment, treated and control cells were processed to obtain gDNA using gDNA isolation kit (Sigma, G1N70) following manufacturer's instructions. Methylated DNA immunoprecipitation (MeDIP) was performed as described before.^{19,70,71} In summary, 5 µg of gDNA was sonicated using Bioruptor sonicator (Diagenode) to a fragment size of approximately 100–400 bp. The sonicated DNA was further heated at 95°C for 10 min and immediately kept in ice. The heated DNA was then incubated with 5-methyl cytosine antibody and normal rabbit IgG antibody at 4°C for overnight. Next day, antibody/DNA complexes were isolated using Dynabeads Protein G (Invitrogen, Cat No. 10004D). The antibody/DNA complex and input DNA were reverse cross-linked in 1% SDS and proteinase K (Invitrogen, 25530049, lot no. 2291039) supplemented in TE buffer for overnight at 65°C. The reverse cross-linked DNA was then purified using PCR purification kit (QIAGEN, 28106). The eluted DNA and 5% input were analyzed by quantitative real-time PCR using the SYBR Green master mix (Promega, A6002, lot no. 0000472956) using primers given in [Table S5](#). Normalization was performed to input using the formula: $[2^{(Ct \text{ input} - Ct \text{ immunoprecipitation (IP)})}]$.⁷² The obtained values were then normalized to the rabbit IgG control IP values for the primer set.

Chromatin immunoprecipitation

ChIP is performed as described previously.^{19,70} In summary, cells were first fixed with 1% formaldehyde and subsequently quenched by 0.125M glycine. Cells were pelleted down and lysed. Next, isolated chromatin was sonicated to an approximate chromatin fragment length of 100–400 bp. The sonicated chromatin was then immunoprecipitated with HIF1α, CTCF and RNA Pol II antibody or corresponding control (immunoglobulin G) antibody at 4°C for overnight. Further, ProteinG dynabeads (Invitrogen, Cat No. 10004D) were added and obtained complex and input DNA were reverse cross-linked with 1% SDS and proteinase K (Invitrogen, 25530049, lot no. 2291039 for overnight) added in TE buffer at 65°C followed by purification using PCR purification kit (QIAGEN, 28106). The immunoprecipitated (IP) DNA and 5% input were analyzed by qRT-PCR using specific primers ([Table S5](#)) and SYBR Green Master Mix (Promega, A6002, lot no. 0000472956). IP DNA values were normalized to input and subsequently normalized to IgG control IP using the following formula: $2^{(Ct \text{ input} - Ct \text{ immunoprecipitation (IP)})}$.⁷² The relative values are represented as mean ± SEM of triplicates.

Chromosome conformation capture (3C) assay

The 3C analysis was performed as described before^{73,74} with slight modifications. In brief, 5×10^6 MCF7 cells were fixed with 1% formaldehyde for 10 min at room temperature, and the reaction was quenched by 0.125 M glycine for 5 min at room temperature. Cross-linked cells were washed with ice-cold PBS, scraped and collected. Cell pellet was re-suspended with 5mL lysis buffer (50 mM Tris-HCl with pH 8.0, 150 mM NaCl, 5 mM EDTA and 0.5% NP-40 with 1× proteinase inhibitor), and incubated on ice for 20 min. Cell pellet was then resuspended in 440ul of Milli-Q water and 60 µL 10×RE buffer (NEB, B7002S) was added, then incubated with 15 µL 10% SDS for 90 min at 37°C with shaking at 900 rpm. Further, 75 µL of 20% Triton X-100 was added and incubated for 90 min at 37°C with shaking at 900 rpm. After heating, 5 µL of aliquot was taken as undigested control. Restriction enzyme was then added sequentially 200 U HindIII (NEB, R3104T, lot No. 10029681) for 4 h at 37°C with shaking at 900 rpm. Again 200 U RE was added and incubated for overnight at 37°C with shaking. At last, final 200U RE was added and incubated for 4 h at 37°C with shaking. 5 µL of aliquot was taken as digested control. After inactivating the restriction enzyme for 15 min at 65°C, fragments were ligated with T4 DNA ligase (NEB, M0202S, lot no. 10113858) overnight at 16°C. Reverse crosslinking was carried out at 65°C for overnight with proteinase K (Invitrogen, 25530049, lot no. 2291039), followed by RNase A (Invitrogen, 12091021) treatment and phenol/chloroform extraction precipitation. The 3C interactions were analyzed by quantitative real-time PCR and product was then run on 2% agarose gel and sequenced. The amount of DNA in the qPCR reactions was normalized across 3C libraries using internal control primers. Primer sequences are listed in [Table S5](#).

Transwell invasion assays

The invasion assay is performed as mentioned previously.^{9,75} In brief, the transwell filter inserts (corning, 3422) were first coated with matrigel (corning, 356230, lot no. 2010001), set for 3 h at 37°C incubator. Further, 3×10^4 cells were suspended in serum free media and were placed in the upper chamber of the transwell setup and in the lower part of the well, 10% FBS supplemented media was added. This setup was incubated for 48–72 h in normoxic or hypoxic conditions. Next, cells in the upper chamber were gently removed from the top of the matrigel and the membrane was fixed in 4% formaldehyde followed by staining with crystal violet (0.05% crystal violet in 10% methanol in 1× PBS). Five random fields were counted using an inverted microscope (Olympus CKX41). All the invasion experiments were performed at least thrice. The % cell invaded is represented as mean ± SEM of triplicates.

3D spheroid invasion assay

3D spheroid invasion assay was performed as described previously.^{76,77} 3×10^3 cells were seeded into an ultra-low attachment (ULA) 96-well U-bottom plate. Spheroids were formed after incubation for 24 h at 37°C, 5% CO2. Collagen I was diluted to the concentration of 1 mg/mL and 100 µl of diluted collagen I solution was added to each well and was kept for 1 h at 37°C to let it polymerize. Spheroid were then transferred into a well of 96-well plate (single spheroids/well) coated with 100 µL 1 mg/mL Collagen I mix., After seeding spheroid into Collagen I mix, the plates were placed into incubator at 37°C, 5% CO2, to set for 1 h. Cell culture medium was added next. MCF7 cells overexpressing an empty vector or COL5A1 isoforms, COL5A1exon64A and COL5A1exon64B spheroids

embedded in a collagen I matrix, allowed to invade for 12 h under normoxic conditions. MCF7 cells transfected with either dCas9-DNMT3A-sgcontrol or dCas9-DNMT3A-sg3 spheroids embedded in a collagen I matrix were allowed to invade for 12 h under hypoxic condition. The images were captured by Thermo Scientific Invitrogen EVOS FL Auto 2 Imaging System and at 10 \times magnifications. The relative invasion area to the initial spheroid area was quantified.

Wound healing and cell counting Kit-8 (CCK-8) viability assays

MCF7 cells (2×10^5) were seeded in 12-well plates and wound healing was performed as before.⁷⁸⁻⁸⁰ After 16 h, the complete medium was replaced with a low concentration of serum fresh medium (2%). Consistently shaped wounds were scratched with a 10- μ L pipette tip across each well at 90% confluence, as described previously.⁸¹ The cells were gently washed with PBS twice to remove loose cells, and serum-free medium was added. To ensure that the wounds with the same wound area were comparable, multiple positioning marks were made at the center of the bared surface. The scratch zones were imaged at 0 h and different time points until the wound is closed. The wound_healing_size_tool (ImageJ plugin) was used for the measurements and to determine the migrating ability of cancer cells.⁸² The migration rate is expressed as the percentage of wound closure as described previously.^{78,83}

Cell proliferation was analyzed using CCK8 assays (CELL COUNTING KIT – 8, Sigma, Cat no. 96992-500TESTS-F) following the manufacturer's protocol. In brief, 8×10^3 MCF7 cells overexpressing the empty vector or COL5A1 isoforms, COL5A1exon64A and COL5A1exon64B individually were seeded into 96 well plates and incubated for different time points in normoxic condition. MCF7 cells transfected with either dCas9-DNMT3A-sgcontrol or dCas9-DNMT3A-sg3 were seeded into 96 well plates (1×10^4 cells/well) and after 12 h, they were kept under hypoxia and after incubation for 24 or 48 h, CCK8 was added. The absorbance was detected at 450 nm after the cells were treated with 10% CCK-8 at 37°C for 2 h.

Western blot

Western blot is performed as described previously.^{9,84} In summary, for protein isolation, urea lysis buffer (8 M urea, 2 M thiourea, 2% CHAPS, 1% DTT) supplemented with 1X Protease inhibitor (leupeptin 10 μ M, pepstatin 5 μ M, EDTA 1mM, AEBSF 200 μ M) was added to the cells washed with 1 \times PBS and spun at 14,000 \times g in 4°C centrifuge. Equal concentration of protein extracts were separated by SDS/PAGE, electroblotted onto PVDF membranes, and were incubated with primary antibodies followed by secondary antibody. Protein bands were visualized using Odyssey Infrared imaging system (Licor). Quantification of the bands was done using ImageJ software.

RNA isolation, cDNA synthesis and quantitative reverse transcription real-time PCR (qRT-PCR)

After treatment, total RNA from cells was extracted using TRIzol (Invitrogen, 15596026) according to the manufacturer's instructions. Further, cDNA was synthesized using PrimeScript 1st strand cDNA Synthesis Kit (TaKaRa, 6110A, lot no. AJX1015N) as per manufacturer's instructions. Next, qPCR was performed with the primer listed in Table S6 with SYBR Green master mix (Promega, 75665) using light cycler 480 II (Roche). The relative gene expression were calculated using the following formula: 2 $^{(Ct_control - Ct_target)}$ where RPS16 is taken as control.⁷² Also, to control gene normalization, exon-level expressions were normalized to a constitutive exon.

RNA interference

Knockdown of the genes were performed as previously described.^{9,85} Briefly, after 24 h of seeding, the cells (MCF7 or HCC1806 cells) were transduced with lentivirus containing small hairpin RNA (shRNA) (Sigma, Mission Human Genome shRNA Library, Table S7) against HIF1A, CTCF, shControl with 8 μ g/mL polybrene (Sigma, H9268) containing media. Further, for 3 days, transduced cells were selected using 1 μ g/mL puromycin (Sigma, P9620). These selected cells were used for downstream experiments. Next, for rescue experiments, CTCF depleted cells were transfected with COL5A1exon64A or COL5A1exon64B transcript overexpression plasmid using Lipofectamine 2000 reagent (Invitrogen, 11668019) as per the manufacturer's instructions.

sgRNA designing and cloning in dCas9-DNMT3A expression vector and cell transfections

GT-Scan web-tool (<http://gt-scan.braembl.org.au/gt-scan/>⁶⁴) and CRISPICK (<https://portals.broadinstitute.org/gppx/crispick/public>⁶⁵) were used to design guide sequences targeting the loci of interest. The designed sgRNA were cloned into dCas9_DNMT3A (Addgene #74407) expression vectors as described previously^{40,86} using Bpil/BbsI (Fermentas, Cat No. ER1012). Additionally, Addgene #71830 plasmid having non-targeting (NT) control guide RNAs (sgcontrol) was used as a control for dCas9-DNMT3A system. The cloned guide RNAs were verified by sequencing using the U6-Forward primer. Sequences of all oligonucleotides are listed in Table S7. For each sgRNA, potential off-target sequences were identified using the off-target site prediction software Cas-OFFinder given in Table S7 (<http://www.rgenome.net/cas-offinder>,⁶³). These primers were used to check the potential off-target sites of dCas9-DNMT3A system and MeDIP-qPCR was carried out at top 2 predicted off-target sites with three or less mismatches compared to the on-target sequence sgRNA. List of primers are given in Table S7. For, dCas9-DNMT3A mediated methylation, MCF7 cells were transfected with the above cloned sgRNA-dCas9_DNMT3A system as described previously.⁹ Briefly, Twenty-four hours post seeding, MCF7 cells were transfected with 3 μ g of sgRNA-dCas9-DNMT3A constructs using 1:3 ratio of DNA to

TurboFect (Thermo scientific, Cst No. R0531) in 60 mm dish and scaled up for large format. Further, cells were then given hypoxia treatment for 24 h and subsequently used for isolation of DNA, RNA, protein or experiments such as ChIP, MeDIP, 3C and invasion assay.

CRISPR-Cas9 mediated disruption of CTCF binding motifs

To edit CTCF motif in *COL5A1* promoter and downstream of exon64A, sgRNA was designed as mentioned above (Table S7). The sgRNAs were cloned into the lentiviral vector pLentiCRISPR-E (Addgene, #78852,⁴¹). The sgRNA sequence was verified by sanger sequencing. MCF7 cells were transduced with lentivirus produced from the sgRNA cloned pLentiCRISPR-E plasmids with pLentiCRISPR-E sgcontrol as the control. Further, the transduced cells were selected using 1 µg/mL puromycin for 3 days. Edited CTCF motif was confirmed by sanger sequencing and ICE analysis.⁶²

GSEA analysis

TCGA tumor samples were stratified according to CTCF-related 193 hypoxic gene expression, as described previously.⁸⁷ Gene set enrichment analysis was performed using GSEA software to check the enrichment of hypoxia pathway genes (using the list from MSigDB,⁸⁸) in 193-high vs. 193-low tumor samples.^{66,89}

COL5A1 expression analysis in publicly available dataset

TCGA breast cancer data was analyzed for *COL5A1* expression levels in normal vs. tumor samples using online data portal GEPIA2.0 (<http://gepia2.cancer-pku.cn/>,⁵⁸). The microarray dataset in Gene Expression Omnibus (GEO) under accession number GSE147516³⁸ was used to analyze the *COL5A1* gene expression and alternative splicing event for *COL5A1* Exon64 in normoxia vs. hypoxia HCC1806 breast cancer cell line. For HTA 2.0 triple-negative breast cancer patient profile (GSE76250³⁹), was analyzed for *COL5A1* expression and *COL5A1*exon64 AS event as before.^{38,87} Splicing events with absolute splicing index (ASI) ≥ 1.5 and $p < 0.05$ were considered to be significant.

COL5A1 promoter region's in silico analysis for transcription factor binding sites and CpG islands

The Eukaryotic Promoter Database was searched for the sequence of the human *COL5A1* gene promoter (<https://epd.vital-it.ch/>,⁴²). Using DBCAT (<http://dbcatt.cgm.ntu.edu.tw/>,⁶¹) and EMBOSS CpGPlot (<https://www.ebi.ac.uk/emboss/cpgplot>,⁶⁰), CpG islands were identified in *COL5A1* promoter (2kb upstream and 0.5kb downstream of TSS, transcript NM_000093). The putative CTCF and HIF1 α binding sites were predicted using JASPAR^{55,56} and CIS-BP.⁵⁹

Retrieval of gene list from various databases

List of gene hallmark of various pathways including EMT was obtained from MSigDB.⁸⁸ Additionally, EMT gene signatures published previously were also retrieved.^{12,13,90–93} Gene overlapping from different sets was performed using online tool <https://bioinformatics.psb.ugent.be/webtools/Venn/>.

In silico analysis of CTCF binding motifs for the COL5A1 locus

We used Find Individual Motif Occurrences (FIMO) software⁵⁴ to search CTCF Binding motifs on *COL5A1* promoter and near exon64 with default parameters. The program generates both the forward and reverse strand hits which are ranked to a logarithmic sequence similarity score on binding locations.

RNA-seq data analysis

MCF7 cells normoxia and hypoxia (16 h) RNA-seq data was obtained from GSE166203.³⁷ Gene expression levels and differentially expressed genes (DEGs) in normoxia vs. hypoxia (16 h) were identified using NOIseq analysis.⁵³ Genes with probability of difference of 0.6 or more were called as differentially expressed genes. Alternative splicing events in hypoxia as compared to the normoxia were retrieved using rMATS (v. 4.0.2).⁵²

ChIP-sequencing

ChIP experiment was performed as above. ChIP-seq library preparation was performed using NEBNext Ultra II DNA library Prep Kit for Illumina (NEB, Cat. E7645S) according to the manufacturer's instructions. 10 ng of input ChIP-enriched DNA was used for ChIP-seq library preparation. Paired-end sequencing (2 × 150 bp) of these libraries was performed in Novaseq X plus platform.

ChIP-seq data analyses

For each library, Reads were trimmed using Trimmomatic (v0.39),⁴⁶ and uniquely aligned to GRCh38 using bowtie2 (v2.7.3a) aligner.⁵⁷ The biological replicates were merged by Samtools v1.9.⁴⁷ MACS2 v2.1.2 was used to call peaks.⁴³ To identify the CTCF differential binding regions in normoxia vs. hypoxia ChIP-seq, DiffBind v3.6.5 was used.⁴⁸ The differential peaks identified by DiffBind were annotated by using BioMart by ENSEMBL.⁴⁹ The CTCF binding sites were considered correlated to gene promoters if present ± 10.0 kb from TSS. The GO analysis was performed using ShinyGO (v 0.80).⁵⁰ Integrative Genomics Viewer was used to visualize the signals.⁴⁵

To find CTCF mediated exon inclusion events under hypoxia, we quantified the distribution of hypoxia-gained CTCF sites (4142) with exons included under hypoxia using BEDtools (v2.26.0)⁵¹ and were put into three groups i) CTCF peak either overlapping with exon or intronic regions downstream of exons +5 kb from the exon 3' end corresponding the roadblock model discussed above¹⁹ ii) CTCF peak present at promoter \pm 10 kb and also in intronic regions upstream of exons +10 kb from the exon 5' end reflecting CTCF-mediated promoter-intragenic looping upstream of exon²⁴ iii) CTCF peak present at promoter \pm 10 kb and also in intronic regions upstream of exon and downstream of exons pertaining roadblock model correlating present study mechanism coupling promoter-exon upstream looping with CTCF mediated RNA Pol II pause at exon. Notably, for the last two models we consider the gene associated with CTCF-mediated splicing if any of the CTCF binding peak (either at promoter, intergenic or downstream of exon) is significantly gain.

CUT&RUN assay

CUT&RUN was performed as stated before.⁹⁴ Briefly, 0.5 million MCF7 cells were fixed with 0.1% formaldehyde for 1 min and quenched with 125 mM glycine. Further cells were scraped, pellet down and nuclei were extracted with nuclear extraction buffer (20 mM HEPES, pH7.9, 10 mM KCl, 0.1% Triton X-100, 20% glycerol, 1 mM MnCl2). The nuclei were washed with wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1x Protease Inhibitor, 0.05% SDS and 1% Triton X-100) and immobilized to Concanavalin-A beads (CST, #93569S), resuspended in 50 μ L antibody buffer (wash buffer, 2 mM EDTA supplemented with 0.01% digitonin) and incubated with 0.5 μ g of the following antibodies: normal rabbit IgG (CST, # 2729S), CTCF (CST, # 3418S) overnight at 4°C. Beads were washed with Cell Permeabilization buffer (Wash Buffer +0.01% digitonin) and incubated with pAG-MNase (CST, #40366S) for 10 min at RT. After activation by CaCl2 beads were incubated for 2 h at 4°C, Stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 μ g/mL RNase A, 50 μ g/mL Glycogen (GlycoBlue Coprecipitant)) was added and incubated at 37°C for 10 min to release CUT&RUN fragments from the insoluble nuclear chromatin. DNA was then decrosslinked overnight at 55°C and purified. Wash buffer, Cell Permeabilization buffer and Antibody buffer were supplemented with 0.05% SDS and 1% Triton X-100. Libraries were prepared using the NEBNext Ultra II DNA library Prep Kit for Illumina (NEB, Cat. E7645S). Two replicates were performed for each condition. Paired-end sequencing (50 bp) was performed on Illumina sequencer. Resulted paired end reads were trimmed using Trimmomatic (v0.39),⁴⁶ concatenated, aligned to hg38 using Bowtie2. To identify the CTCF differential binding regions in normoxia vs. hypoxia CTCF CUT&RUN assay, DiffBind v3.6.5 was used.⁴⁸ The differential peaks identified by DiffBind were annotated by using BioMart by ENSEMBL⁴⁹ and EDP_new database.⁴⁴ The CTCF binding sites were considered correlated to gene promoters if present \pm 5.0 kb from TSS. The GO analysis was performed using ShinyGO (v 0.80).⁵⁰ Integrative Genomics Viewer was used to visualize the signals.⁴⁵

Circular chromosome conformation capture (4C)

4C-seq was performed as described before in van de Werken et al., 2012.⁷⁴ In brief, MCF7 shControl cells and CTCF depleted cells were kept under normoxia and hypoxia for 24 h and then 4C-seq was performed using NlalII as primary restriction enzyme and DpnII as secondary restriction enzyme. The resulted 4C-seq libraries were PCR amplified using the COL5A1 promoter primers designed for the viewpoint (Table S5). The PCR products were purified libraries for sequences were made using NEBNext Ultra II DNA library Prep Kit for Illumina (NEB, Cat. E7645S) and subjected to next-generation sequencing with Illumina using 100 bp paired-end reads. The sequenced reads were aligned to hg38 assembly using pipe4C.⁶⁸ 4Cseqpipe was used to process the sequenced data.⁶⁷ 4C-seq images were generated using linear mean at a 5kb resolution.

QUANTIFICATION AND STATISTICAL ANALYSIS

The data is reported as mean \pm SEM. Each experiment involved at least three independent biological replicates. All statistical analyses were carried out using the GraphPad Prism 8 software. The figure legends reflect the statistical tests used and the specific p values.