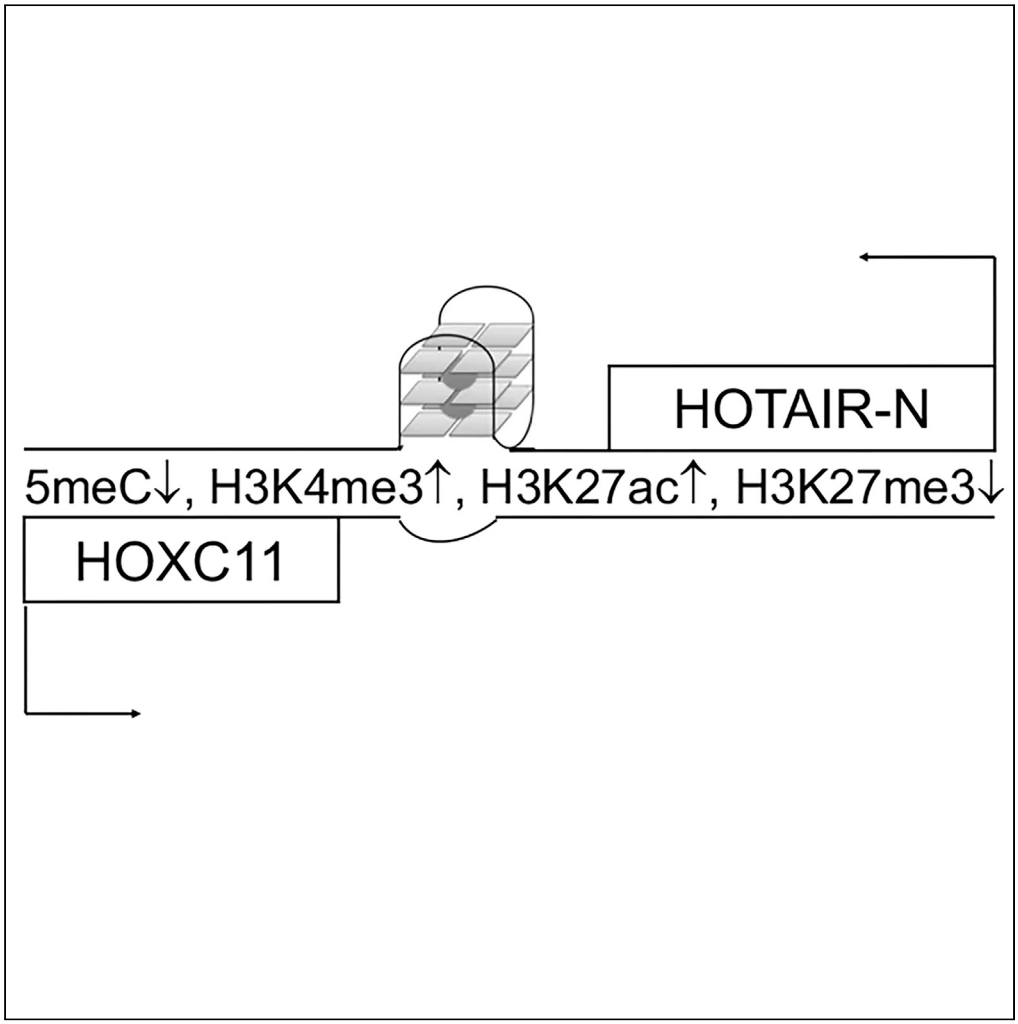
iScience

CellPress
OPEN ACCESS

Article

G-quadruplex is critical to epigenetic activation of the lncRNA HOTAIR in cancer cells



Xiaohan Qu, Zhen Lin, Janarthanan Jayawickramarajah, ..., Fang Fang, Shankar Balasubramanian, Bin Shan

han_seal@163.com

Highlights

HOTAIR-N that overlaps with HOXC11 is the major isoform in cancer cells

HOTAIR-N is required for expression of the cancer-related transcriptomes

G-quadraplex in the HOTAIR-N CpG island is critical to upregulation of HOTAIR-N

Qu et al., iScience *26*, 108559
December 15, 2023
https://doi.org/10.1016/j.isci.2023.108559

iScience

CellPress
OPEN ACCESS

Article

G-quadruplex is critical to epigenetic activation of the lncRNA HOTAIR in cancer cells

Xiaohan Qu,[1,8,*] Zhen Lin,[2] Janarthanan Jayawickramarajah,[3] John S. Alsager,[4] Emily Schmidt,[3] Kenneth P. Nephew,[5] Fang Fang,[5] Shankar Balasubramanian,[6,7] and Bin Shan[4]

SUMMARY

The cancer-promoting lncRNA HOTAIR has multiple isoforms. Which isoform of HOTAIR accounts for its expression and functions in cancer is unknown. Unlike HOTAIR's canonical intergenic isoform NR_003716 (HOTAIR-C), the novel isoform NR_047517 (HOTAIR-N) forms an overlapping antisense transcription locus with HOXC11. We identified HOTAIR-N as the dominant isoform that regulates the gene expression programs and networks for cell proliferation, survival, and death in cancer cells. The CpG island in the HOTAIR-N promoter was marked with epigenetic markers for active transcription. We identified a G-quadruplex (G4) motif rich region in the HOTAIR-N CpG island. Our findings indicate that G4s in HOTAIR-N CpG island is critical for expression of HOTAIR-N in cancer cells. Disruption of G4 may represent a novel therapeutic approach for cancer. The transcriptomes regulated by HOTAIR-N and Bloom in cancer cells as provided herein are important resources for the exploration of lncRNA, DNA helicases, and G4 in cancer.

INTRODUCTION

Long noncoding RNAs (lncRNA) have emerged as a new category of tumor suppressor and oncogenes in cancer. Recent advances in detection of lncRNAs have enabled fast and reliable discoveries of these lncRNA in cancer cells and tumor tissues.[1–3] HOTAIR is a prime example of *trans*-acting lncRNA that promote cancer.[4,5] HOTAIR is located in the homeobox gene C cluster (HOXC) and assigned with three isoforms in RefSeq since the initial discovery of HOTAIR's canonical transcript NR_003716 (named HOTAIR-C hereafter) (Figure 1A).[6,7] HOTAIR-C is intergenic and located between two protein-coding genes HOXC11 and HOXC12. It remains unclear how each isoform contributes to the expression and functions of HOTAIR in cancer. The current paradigm is that HOTAIR-C accounts for elevated expression of HOTAIR in cancer although the methods employed in the previous studies were not able to distinguish contribution from each isoform.[4,8–10]

Among all three isoforms of HOTAIR in the latest release of RefSeq, we are intrigued by NR_047517 that has not been characterized previously (Figure 1A).[6,7] We named this isoform HOTAIR-N to distinguish it from two other intergenic isoforms, HOTAIR-C and another uncharacterized isoform NR_047518 (named HOTAIR-U hereafter). HOTAIR-N is transcribed from the first intron of HOXC11 in an antisense fashion and the overlapping region contains a strong CpG island with 160 CpG sites, named the HOTAIR-N CpG island hereafter (Figure 1A).[11] This configuration resembles other gene loci in the HOXC cluster that are usually enriched with CpG islands, overlapping sense-antisense gene pairs, and G-quadruplex (G4) forming motifs.[7,12–14] G4s are widespread in the human genome and reported to regulate genome stability and transcription in physiological and pathological conditions.[14] Given the genomic configuration of HOTAIR-N, it is conceivable that regulation of HOTAIR-N in cancer cells involves cytosine methylation to form 5-methylcytosine (5meC) on CpG sites, histone modifications, and G4 motifs.

In the current study, we aim to determine whether HOTAIR-N is upregulated in cancer cells and to elucidate the epigenetic mechanisms that mediate upregulation of HOTAIR-N in cancer cells. We demonstrate that HOTAIR-N is the dominant isoform with respect to expression and functions of HOTAIR in cancer cells. Moreover, epigenetic elements within the HOTAIR-HOXC11 CpG island, especially G4 formation, are critical for upregulation of HOTAIR-N expression.

[1]Department of Thoracic Surgery, The First Hospital of China Medical University, Shenyang, Liaoning 110001, China
[2]Deparmtent of Pathology, Tulane University School of Medicine, New Orleans, LA 70112, USA
[3]Department of Chemistry, Tulane University, New Orleans, LA 70118, USA
[4]Department of Biomedical Sciences, Elson S Floyd College of Medicine, Washington State University, Spokane, WA 99202, USA
[5]Medical Sciences, Cell and Molecular Cancer Biology Program, Indiana University School of Medicine, Bloomington, IN 47405, USA
[6]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK
[7]Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK
[8]Lead contact
*Correspondence: han_seal@163.com
https://doi.org/10.1016/j.isci.2023.108559

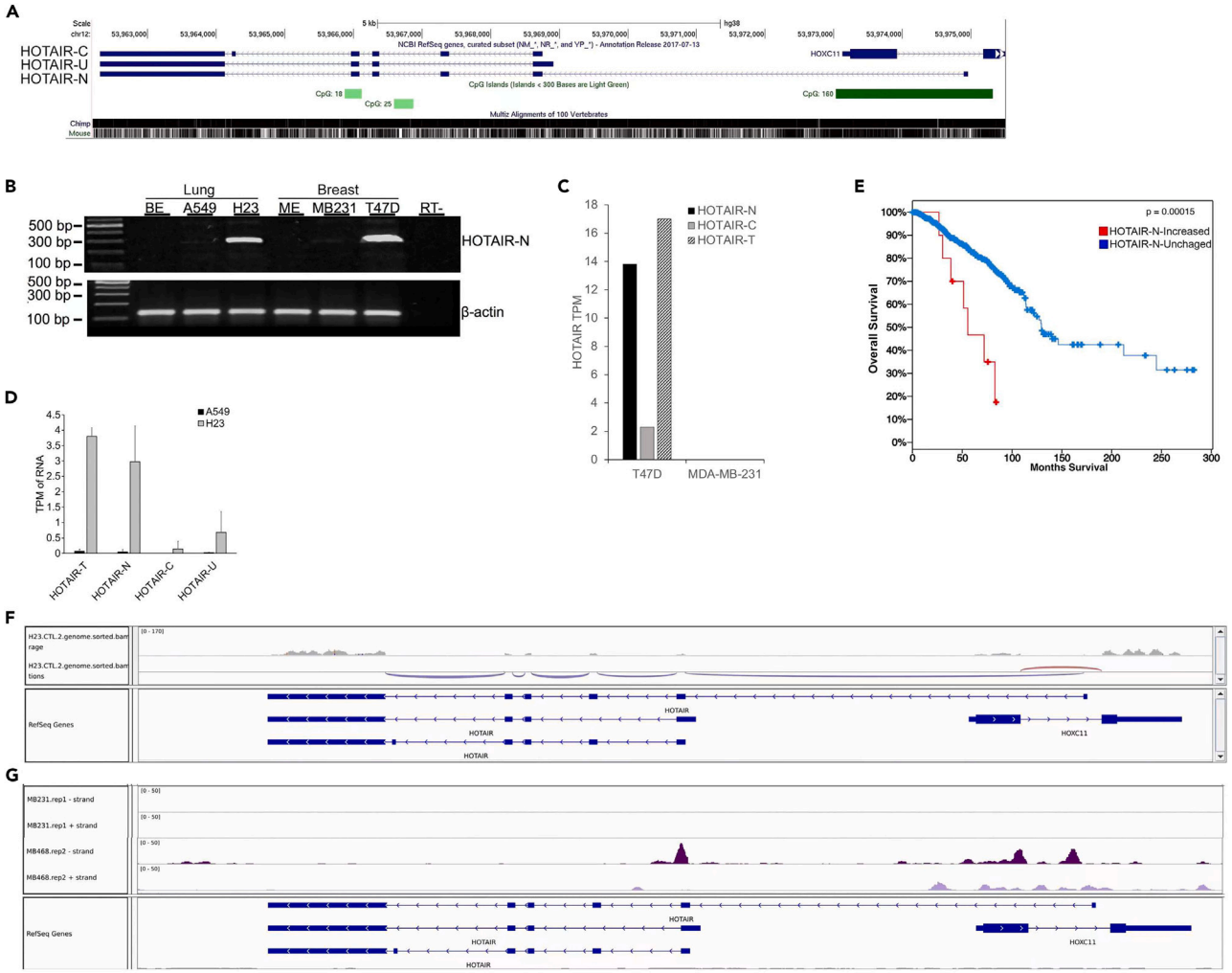


Figure 1. Elevated expression of HOTAIR-N in cancer cells

(A) A screenshot of the human HOTAIR isoforms in RefSeq using the UCSC Genome Browser. HOTAIR-N corresponds to NR_047517. HOTAIR-C corresponds to NR_003716. HOTAIR-U corresponds to NR_047518.

(B) Total cell RNA was extracted from the indicated cell lines (BE, normal bronchial epithelial cell line; ME, normal mammary epithelial cells; RT-, negative control of reverse transcription). The RNA levels of HOTAIR-N and the housekeeping gene β-actin were assessed using RT-PCR and visualized using SybrGold staining on agarose gels. The image was representative of three independent experiments.

(C) The amount of total HOTAIR (HOTAIR-T), HOTAIR-N, and HOTAIR-C was compared between MDA-MB-231 and T47D cells in the RNA-SEQ data deposited on NCBI SRA. Each transcript was quantified as transcript per million (TPM) using RSEM.

(D) RNA-SEQ was carried out on total RNA extracted from A549 and H23 cells. Total HOTAIR (HOTAIR-T) and each isoform was quantified as TPM using RSEM. The means were obtained from three biological replicates.

(E) Overall survival was compared between the patients with HOTAIR-N increased in the tumor over the matching tumor tissues and the rest patients in the TCGA Invasive Breast Carcinoma cohort. We used cBioportal to carry out Kaplan-Meier Estimate and log-rank test (cBioportal.org).

(F) The sequencing reads and splicing junctions of a representative RNA-SEQ of H23 cells were illustrated on IGV.

(G) The sequencing reads of GRO-SEQ of MDA-MB-231 and MDA-MB-468 cells were illustrated in a strand specific fashion on IGV.

RESULTS

Elevated expression of HOTAIR-N in cancer cells

Addition of the HOTAIR-N-specific first exon (59 bp) did not result in protein-coding potential as predicted by ORFfinder (www.ncbi.nlm.nih.gov/orffinder/) (Figure S1). We designed a pair of HOTAIR-N-specific primers that spans the first and third exons of HOTAIR-N (+32 to +315 relative to HOTAIR-N's transcription initiation site) and examine HOTAIR-N expression using RT-PCR in two types of non-cancer cells, the immortalized HBEC3 (lung) and human mammary epithelial cells (HMEC) (breast), and four tissue type matching cancer cell lines were selected, two lung cancer cell lines A549 and H23, and two breast cancer cell lines MDA-MB231 and T47D.[15] We observed undetectable

to modest expression of HOTAIR-N in HBEC3, A549, HMEC, and MDA-MB-231 cells (Figure 1B). In contrast, H23 and T47D cells exhibited robust expression of HOTAIR-N (Figure 1B).

To determine relative contribution of HOTAIR-C and HOTAIR-N to the expression of total HOTAIR in cancer cells, we examined the RNA-SEQ data of MDA-MB-231 and T47D cells available at NCBI Sequence Read Archive (SRP005601).[16] Consistent with our RT-PCR results T47D exhibited higher expression of HOTAIR, including HOTAIR-N, HOTAIR-C, and total HOTAIR than MDA-MB-231 cells (Figure 1C). HOTAIR-N accounted for ~75% of total HOTAIR in T47D cells (Figure 1C). We further examined expression of the HOTAIR isoforms in A549 and H23 cells using RNA-SEQ (Figure 1D). The RNA levels of total HOTAIR and HOTAIR-N in H23 cells was 57-fold and 63-fold higher than that in A549 cells, respectively. HOTAIR-N accounted for 78% of total HOTAIR in H23 cells. The RNA levels of HOTAIR-C and HOTAIR-U in H23 cells, albeit also higher, accounted for only 22% when combined.

To determine how much HOTAIR-N accounts for the expression of HOTAIR in human tumor tissues we analyzed The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) and breast invasive carcinoma (BRCA) RNA-SEQ data using the RSEM algorithm that is tailored for quantification of isoforms.[17,18] In each dataset we selected the paired tumor and non-tumor samples that exhibit elevated expression of HOTAIR in tumor tissues over their paired non-tumor tissues. In the invasive breast carcinoma dataset, we screened the paired tumor and nontumor tissues and identified 14 paired samples that exhibited higher increase in expression of total HOTAIR in tumor over their paired non-tumor tissues. In 10 of those 14 tumor samples, HOTAIR-N exhibited the highest expression among all three isoforms and accounted for ~57% of HOTAIR in breast cancer as we previously reported.[19] Similarly, in 12 selected pairs of matching tumor and non-tumor lung tissues HOTAIR-N was the highest expressed isoform in 10 tumor samples and accounted for ~72% of total HOTAIR expression on average (Table S2). In contrast, HOTAIR-C was a minor isoform in both datasets as it accounted for ~26% and ~31% on average in lung and breast, respectively (Table S2).[19] Similar findings were observed in the paired tumor and nontumor samples from the TCGA Kidney Renal Papillary Cell Carcinoma (KRPCC) cohort (Table S3). Moreover, we compared overall survival of the breast cancer samples with increased HOTAIR-N over the paired normal samples with the rest samples from the invasive breast carcinoma cohort of the TCGA project. We used cBioportal platform to carry out Kaplan-Meier Estimate and log-rank test (cBioportal.org). Indeed, our analysis suggested an association between elevated expression of HOTAIR-N and was associated with poor overall survival in the TCGA Invasive Breast Carcinoma cohort (Figure 1E). Due the small sample size of the HOTAIR-N expression elevated group, this association needs to be further validated in future studies that include sufficient samples for more rigorous statistical analyses.

To confirm the transcription initiation site of HOTAIR-N as specified in RefSeq, we carried out 5'RACE on RNA extracted from H23 cells using a primer complementary to a fragment in the last exon that is common to all three HOTAIR isoforms (Figure 1A). The 5'RACE clones were sequenced and the sequencing results were blasted using BLAT.[20] The transcription initiation of HOTAIR-N was mapped precisely at the site as specified by RefSeq (Figure S2).[6] To further confirm expression of the HOTAIR-N transcript, we examined the transcription initiation and splicing of HOTAIR-N in H23 cells based on the RNA-SEQ data. Indeed, the HOTAIR-N transcript was initiated and spliced from the 1st intron of HOXC11 in an antisense fashion (Figure 1F). To capture active transcription from HOTAIR-N, we analyzed the RNA-SEQ and GRO-SEQ data of two human basal-like breast cancer cell lines, MDA-MB-231 and MDA-MB-468 (GSE96867).[21] The RNA levels of total HOTAIR and HOTAIR-N in MDA-MB-468 cells were 283- and 406-fold of that in MDA-MB-231 cells, respectively (Table S4). Moreover, HOTAIR-N accounted for 70% of total HOTAIR in MDA-MB-468 cells (Table S4). GRO-SEQ is able to directly measure nascent RNA production.[22] We observed robust nascent RNA production through the HOTAIR-N region that overlaps with HOXC11 and harbors the CpG island (hg/38/human chr12:53973032-53975319) (Figure 1G). The HOTAIR-N CpG island yielded an average of 387 reads in MDA-MB-468 cells compared to 0 read in MDA-MB-231 cells (Table S5).

Requirement of HOTAIR-N for the gene expression programs related to cancer

To confirm that HOTAIR-N is a major functional isoform in cancer cells, we transfected the HOTAIR-N-specific siRNAs (HOTAIR-Nsi) into HOTAIR-intermediate H23 and HOTAIR-high T47D cells. The RNA levels of total HOTAIR and HOTAIR-N exhibited a ~70% reduction by HOTAIR-Nsi relative to the control siRNA (CTLsi) in H23 and T47D cells (Figures 2A and 2B). HOTAIR-Nsi diminished proliferation of T47D cells by ~2.3-fold because the HOTAIRsi group exhibited only 2.9-fold increase in cell count at 72 h after transfection vs. robust a 7-fold increase in the CTLsi group (Figure 2C, $p < 0.001$). Moreover, the HOTAIR-Nsi transfected T47D cells exhibited a 61% reduction in cell viability when compared to the CTLsi group as measured by MTT assays (Figure 2D, $p < 0.01$).

We then carried out RNA-SEQ to gain insight into how HOTAIR-N regulates proliferation and viability. HOTAIR-Nsi significantly altered the expression of 1829 genes in T47D cells after filtering with a fold change greater than 2 and an false discovery rate (FDR) smaller than 0.01 (Figures 2E; Table S6). Similarly, 904 genes were differentially expressed between HOTAIR-Nsi and CTLsi transfected H23 cells (Table S7). It was noteworthy that HOTAIR-Nsi substantially reduced the expression of its partner EZH2 to 38% relative to CTLsi in T47D cells (Table S6). We determined the pathways and networks that were altered by HOTAIR-Nsi using Ingenuity Pathway as we previously described.[23] HOTAIR-Nsi significantly altered 259 canonical pathways in T47D cells (Table S8). The top 50 canonical pathways altered by HOTAIR-Nsi in T47D cells featured cancer related pathways, such as Molecular Mechanisms of Cancer, ERK/MAPK Signaling, STAT3 Pathway, NF-κB Signaling, and CDK5 Signaling (Figure 2F, for details see Table S8). HOTAIR-Nsi altered similar gene expression programs and canonical pathways in H23 cells (Figures S3A and S3B; Table S9). HOTAIR-Nsi altered 25 networks in T47D cells that were exemplified by the networks that regulate cell cycle, cell death, and survival (Figure 2G; Table S10). HOTAIR-Nsi altered similar networks in H23 cells (Figure S3; Table S11). GO term analyses further confirmed a critical role of HOTAIR-N as 11 GO terms altered by HOTAIR-N in T47D cells and H23 cells

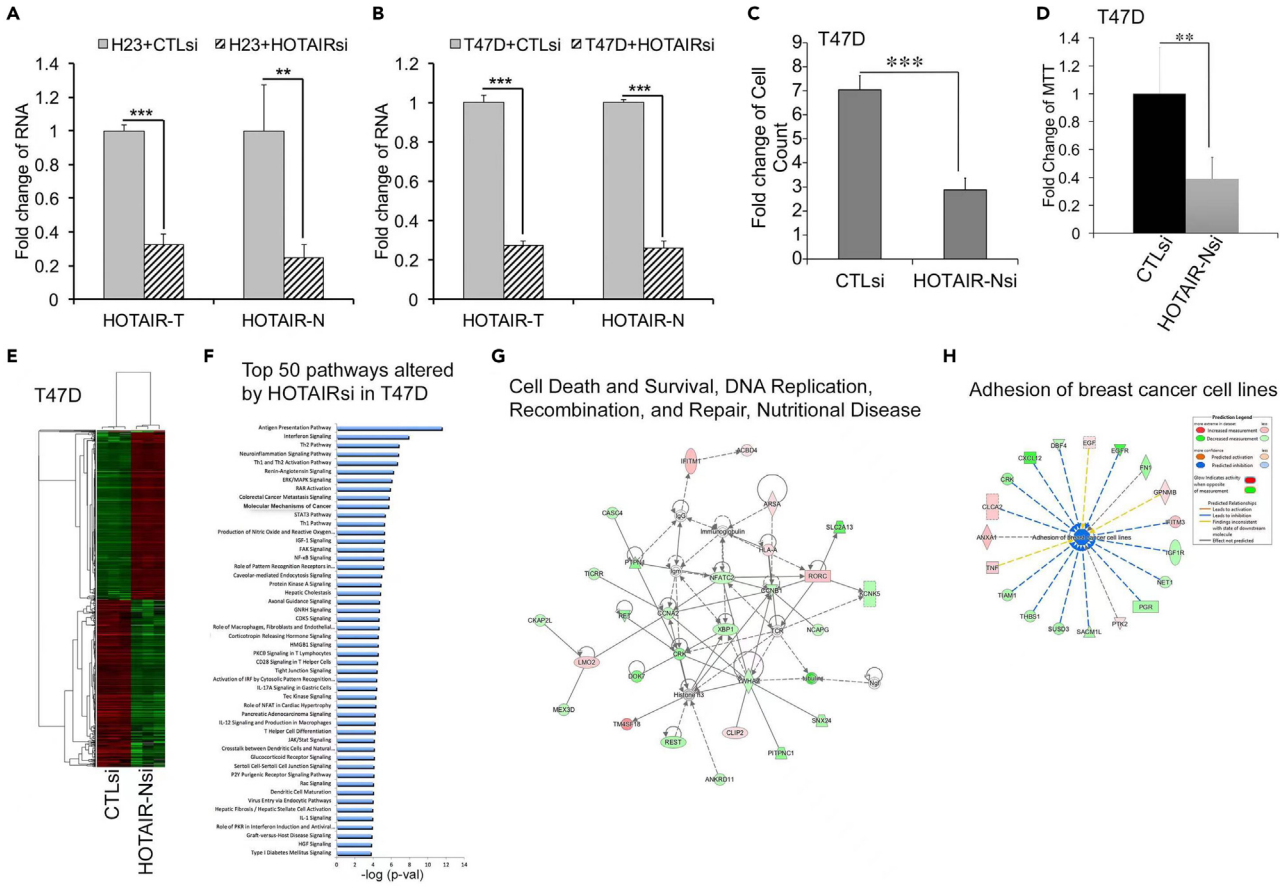


Figure 2. Effects of HOTAIR-N knockdown on cancer cells

(A) The HOTAIR-N-specific siRNAs (HOTAIR-Nsi) were transfected into H23 cells. Total RNA was extracted and RNA levels of total HOTAIR (HOTAIR-T) and HOTAIR-N were assessed using real-time qPCR. A fold change was obtained by normalizing to the house keeping gene GAPDH and setting the values from the control siRNA (CTLsi) transfected group to one.

(B) Similar to part A except that T47D cells were transfected with the indicated siRNAs.

(C) Similar to part B except that cell proliferation was measured in T47D cells.

(D) Similar to part B except that cell viability was measured using MTT assays in T47D cells. A fold change of the MTT values was obtained by setting the values from the CTLsi group to one.

(E) T47D cells were transfected with either CTLsi or HOTAIR-Nsi. Total cell RNA were extracted and processed for RNA-SEQ. The differentially expressed gene set between CTLsi and HOTAIR-Nsi was established from RNA-SEQs of the triplicates of each group. The differentially expressed gene sets was illustrated in a heatmap.

(F) Ingenuity Pathway was employed to analyze the differentially expressed gene set between the CTLsi and HOTAIR-Nsi transfected T47D cells (part E) and to identify the pathways altered by HOTAIR-Nsi. The 50 most significantly altered pathways by HOTAIR-Nsi in T47D cells were illustrated after filtering with a fold change >2 and an FDR <0.01.

(G) Ingenuity Pathway was employed to identify the cancer related networks that were significantly altered by HOTAIR-Nsi in T47D cells through analysis of the RNA-SEQ data as seen in part E. The representative altered network Cell Death and Survival was illustrated using Ingenuity Pathway.

(H) Gene Ontology (GO) analyses was carried out on the RNA-SEQ data as seen in part E to identify the GO terms that were significantly altered by HOTAIRsi in T47D cells. The representative altered GO terms enrichment of Adhesion of Breast Cancer Cell Lines was illustrated. Mean and standard deviations were obtained from at least three independent experiments. ** and *** indicate a p value <0.01, 0.001, respectively.

were directly related to breast cancer, including proliferation, apoptosis, survival, adhesion, migration, and invasion (Figures 2H; Table S12; Figure S3C).

To further examine the transcriptome regulated by HOTAIR-N, HOTAIR-N was knockdown using a different siRNA that targets all three isoforms in a third cancer cell line, MCF-7-TNR that exhibit high expression of HOTAIR.[24] RNA-SEQ was employed to examine the RNA levels of each HOTAIR isoform. HOTAIR-N accounted for 75% of total HOTAIR, whereas HOTAIR-C and HOTAIR-U accounted for 9% and 16%, respectively (Figure S4A). The HOTAIR-specific siRNA reduced all three isoforms by greater than 70%. Moreover, the expression of 545 genes were altered by HOTAIRsi (Table S13). The cancer-related pathways and networks were also altered by HOTAIR knockdown in MCF-7-TNR cells (Tables S14 and S15; Figures S4B and S4C).

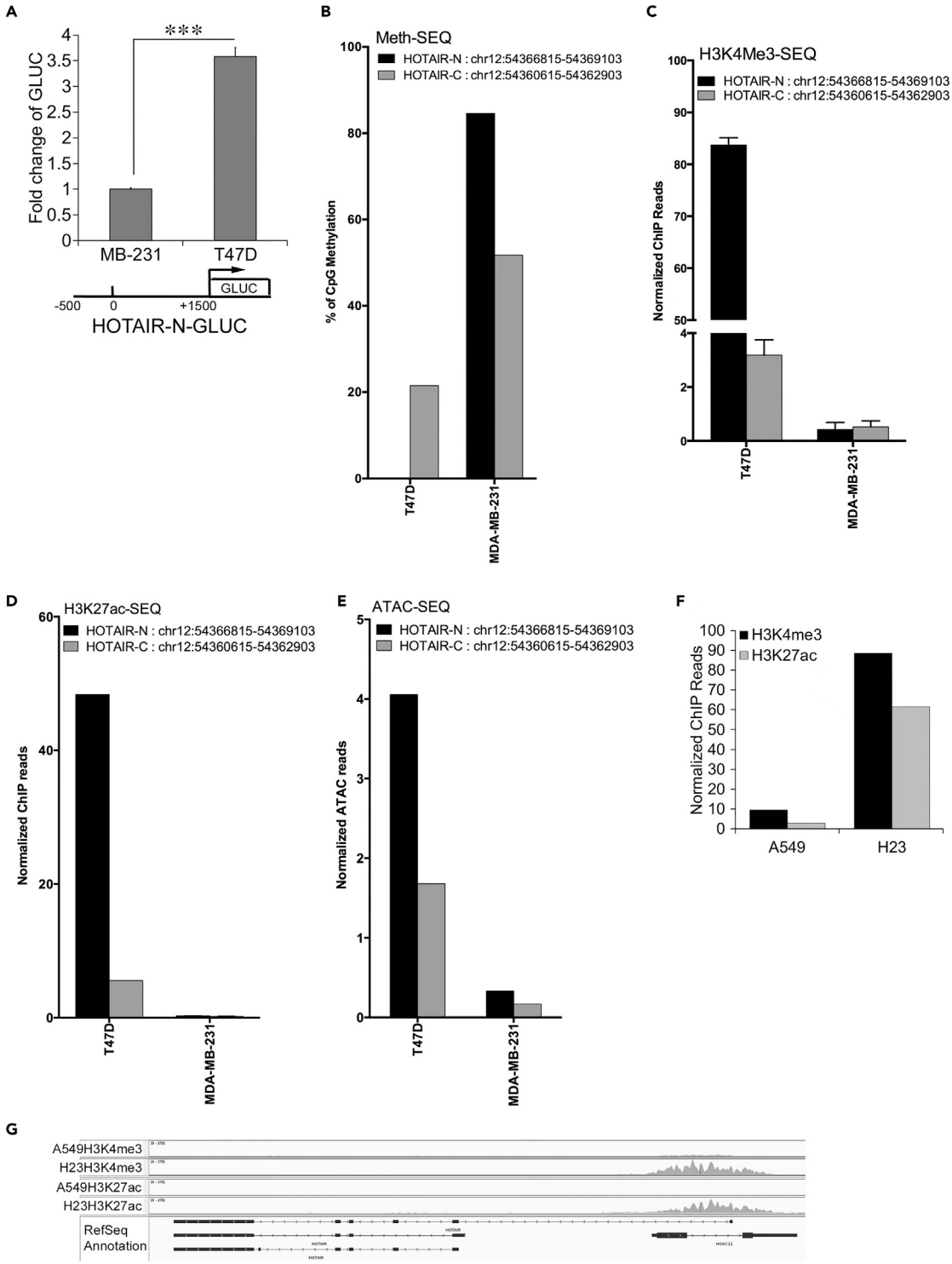


Figure 3. Association between epigenetic markers and the expression of HOTAIR-N

(A) The HOTAIR-N-GLUC and SEAP reporters were transfected into T47D and MDA-MB-231 cells. The ratios of HOTAIR-N-GLUC over SEAP were compared between T47D and MDA-MB-231 cells. A fold change of the ratios was obtained by setting the values from MDA-MB-231 cells to one. Mean and standard deviations were obtained from three independent experiments. *** indicates a p value <0.001.

Figure 3. *Continued*

(B) Abundance of 5meC in the HOTAIR-N CpG island and the corresponding region in the HOTAIR-C locus was quantified by extracting and analyzing genome wide GpG methylation sequencing (Meth-SEQ) obtained from MDA-MB-231 and T47D cells. Percentage of the methylated CpG sites in HOTAIR-N CpG island and the HOTAIR-C corresponding region was compared between MDA-MB-231 and T47D cells.

(C) Similar to part B except that the H3K4me3 bound HOTAIR-N CpG island and the HOTAIR-C corresponding region were compared between the two cell lines.

(D) Similar to part B except that the H3K27ac bound HOTAIR-N CpG island and the HOTAIR-C corresponding region were compared between the two cell lines.

(E) Similar to part B except that ATAC-SEQ reads were compared between the two cell lines.

(F) Similar to part B except the H3K27ac and H3K4me3 bound HOTAIR-N and HOTAIR-C were compared between A549 and H23.

(G) The ChIP-SEQ reads for the H3K27ac and H3K4me3 bound HOTAIR-N CpG island and the HOTAIR-C corresponding region were illustrated on the genome browser.

Gene Set Enrichment Analysis (GSEA) analyses was employed to gain insight into potential mechanisms underlying the biological effects of HOTAIR-Nsi. A total of 1006 gene sets in T47D cells were enriched when the GSEA-recommended cutoff of FDR smaller than 0.25 was applied (Table S16). Consistent with the previous findings the EZH2 target and H3K27me3 gene sets were enriched in the panel (Table S16). Moreover, the DNA methylation gene sets were also enriched (Table S16). Similarly, the same epigenetic gene sets were also enriched in MCF-7-TNR when HOTAIR-N was knocked down (Table S17).

Chromatin stage of the HOTAIR-N CpG island in cancer cells

We were intrigued by the HOTAIR-N CpG island (−363 to +1925 relative to HOTAIR-N's transcription initiation site) that is actively transcribed as revealed by our GRO-SEQ analyses (Figure 1G). Thus, we set to determine whether the HOTAIR-N CpG island region is more transcriptionally active in the HOTAIR-N high T47D cells than that in HOTAIR-N low MDA-MB-231 cells. We transfected a reporter HOTAIR-N-GLUC that is controlled by the HOTAIR-N CpG island region (−500 to +1500 relative to the HOTAIR-N transcription initiation site). As expected, HOTAIR-N-GLUC in T47D cells was 3.4-fold higher than that in MDA-MB-231 cells (Figure 3A).

We compared the percentage of methylated cytosine in the CpG island between MDA-MB-231 and T47D cells by analyzing genome-wide bisulfite treatment sequencing data (BS-SEQ) (SRA# SRP005601).[16] Indeed, 84.6% of the CpG sites were methylated in MDA-MB-231 cells, whereas only 0.1% of those CpG sites were methylated in T47D cells (Figure 3B). In contrast, cytosine methylation in the HOTAIR-C corresponding region (−363 to +1925 relative to HOTAIR-C's transcription initiation site) exhibited a modest difference between MDA-MB-231 (51.7%) and T47D (21.5%) cells (Figure 3B). Trimethylation of histone H3 lysine 4 (H3K4me3) in gene promoters is a hallmark of actively transcribed lncRNAs.[25] Hence, we surveyed H3K4me3 ChIP-SEQ data of MDA-MB-231 and T47D cells (SRA# SRP069855 and SRP073742).[26,27] We observed low abundance of H3K4me3 in the HOTAIR-N CpG island in MDA-MB-231 cells and high abundance of H3K4me3 in T47D cells (Figure 3C). On the other hand, we observed minimal amount of H3K4me3-associated HOTAIR-C corresponding region (−363 to +1925 relative to HOTAIR-C's transcription initiation site of HOTAIR-C) in both cell lines (Figure 3C). Acetylation of histone H3 lysine 27 (H3K27ac) is another hallmark of actively transcribed genes.[28] We surveyed H3K27ac ChIP-SEQ data of MDA-MB-231 and T47D cells (SRA# SRP028597 and SRP052748).[29,30] We observed low abundance of H3K27ac in the HOTAIR-N CpG island in MDA-MB-231 cells and high abundance in T47D cells (Figure 3D). In contrast, minimal amount of H3K27ac was observed in the HOTAIR-C corresponding region (Figure 3D). Open chromatin is associated with active transcription and successfully assessed globally by an assay for transposase-accessible chromatin using high throughput sequencing (ATAC-SEQ).[31] We surveyed ATAC-SEQ and identified more transposase-accessible reads of the HOTAIR-N CpG island in T47D cells than MDA-MB-231 cells (SRA# SRP062544 and SRP078834) (Figure 3E). Similar correlation between epigenetic codes and the expression of HOTAIR-N was also observed in human lung cancer cell lines. H3K4me3 and H3K27ac exhibited strong association with the HOTAIR-N CpG island only in the HOTAIR-intermediate H23 cells, but not HOTAIR-low A549 cells (Figures 3F and 3G) (GSE104481, GSE29611).[32,33] Moreover, the HOTAIR-C corresponding region was devoid of either histone modifications in both cell lines (Figures 3F and 3G). To further elucidate the role of cytosine methylation and histone acetylation in regulation of the HOTAIR-N expression, we chose to inhibit DNA methyltransferase (DNMT) and histone deacetylase, and then examined the expression of HOTAIR-N. The RNA levels of total HOTAIR, HOTAIR-N, and HOXC11 were substantially increased in the HOTAIR-N low A549 and MDA-MB-231 cells upon exposure to 5-Aza-2dC (5 μM), an inhibitor of DNMT for 72 h (Figures 4A and 4B). Similarly, exposure to TSA (250 nM, 48 h), an inhibitor of HDAC, increased the RNA levels of total HOTAIR, HOTAIR-N, and HOXC11 (Figures 4C and 4D).

G4 motifs in the HOTAIR-N CpG island

CpG islands are frequently enriched with G4s.[34] Thus, we searched for G4 motifs in the HOTAIR-N CpG island using QGRS Mapper, an algorithm that identifies G4 forming sequences.[35] We identified a ~500 bp G4 dense region in the HOTAIR-N CpG island that harbors 5 G4 motifs using the formula [G3N1–7G3N1–7 G3N1–7 G3] (Table S18; Figure S5). We selected the two highest scoring G4 motifs (chr12:54368091-54368115 with a score of 60, named G4-60; and chr12:54,367,915-54,367,931 with a score of 41, named G4-41) for further analyses. We used circular dichroism (CD) spectroscopy, a standard *in vitro* G4 structure determination assay, to analyze the structures formed by oligonucleotides G4-60 and G4-41.[36,37] The CD spectra of both G4-60 and G4-41 (10 μM) displayed a trough at 240 nm and a positive peak at 260 nm in 10 mM Tris, 80 mM KCl buffer (pH 7.5) (Figures 5A and 5B). This type of CD signature is indicative of formation of a parallel G4.[36,37] Moreover, the UV-Vis absorption spectra of TMPyP4, a commonly used chromophore that binds to G4, exhibited a characteristic shift to longer wavelengths in combination with hypochromicity, upon addition of increasing concentrations of G4-60 or G4-41 (Figures 5C and 5D).[38,39] The HOTAIR-N CpG island was enriched with G4 SEQ reads and peaks as illustrated in Figure 5E. Similarly,

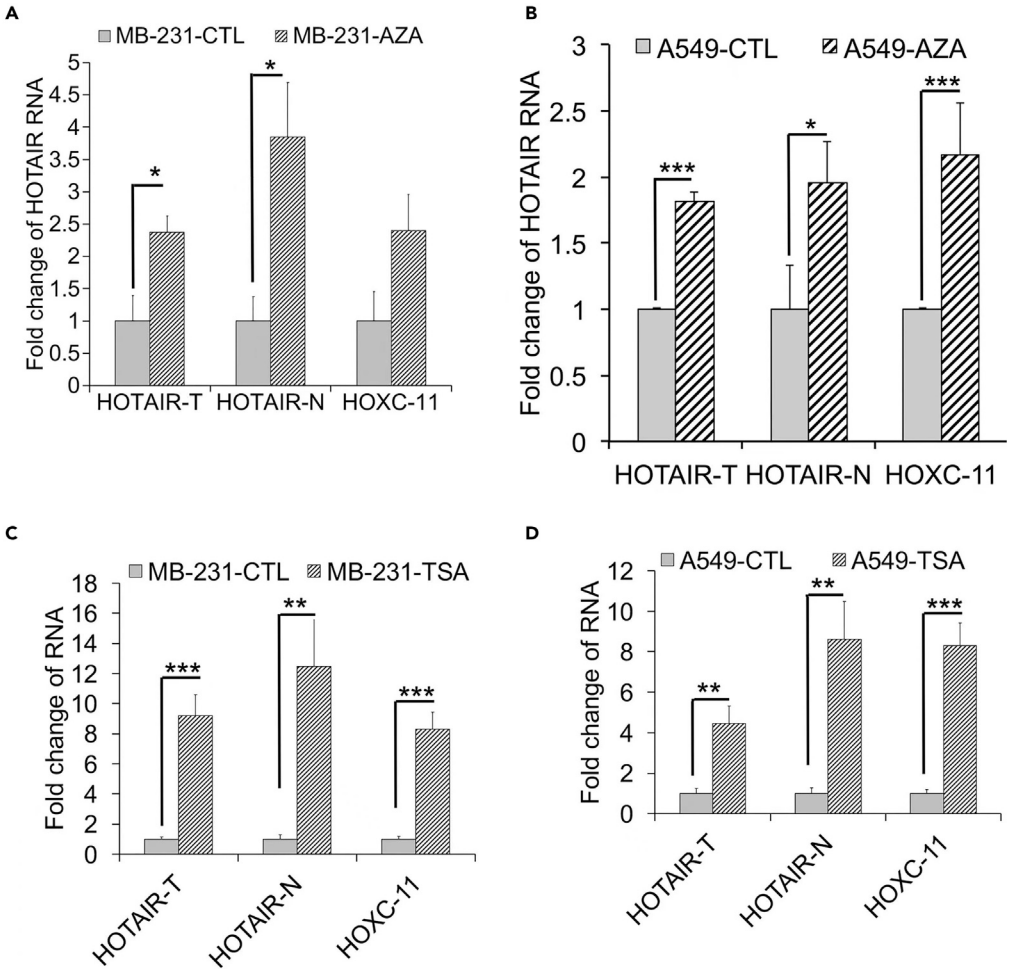


Figure 4. Regulation of the expression of HOTAIR-N by DNMT and HDAC

(A) MDA-MB-231 cells were exposed to the DNMT inhibitor 5′AZA (5 μM) for 72 h. Total RNA was extracted from the treated cells and assessed for the RNA levels of HOTAIR-T, HOTAIR-N, and HOXC11 using real-time qPCR. A fold change was obtained by normalizing to the house keeping gene GAPDH and setting the values from the DMSO group to one (CTL).

(B) Similar to part A except that A549 cells were exposed to 5′AZA (5 μM) for 72 h.

(C) Similar to part A except that MDA-MB-231 cells were exposed to the HDAC inhibitor TSA (250 nM) for 48 h.

(D) Similar to part C except that A549 cells were exposed to TSA. Mean and standard deviations were obtained from three independent experiments. *, **, and *** indicate a p value <0.05 and 0.001, respectively.

the BG4 ChIP reads were enriched in the HOTAIR-N CpG island in the immortalized keratinocytes HACAT (Figure 5F) and the BG4 ChIP signals were further increased by Entinostat, an inhibitor of histone deacetylase (Figure 5F).[13] Moreover, our analysis of the paired RNA-SEQ datasets (SRA# SRP068243) revealed that Entinostat increased the expression of HOTAIR-N and total HOTAIR in HACAT cells (Figure 5G).

G4-mediated expression of HOTAIR-N

To assess the role of G4 in regulation of the expression of HOTAIR-N, we treated MDA-MB231, A549 cells, and H23 cells with Pyridostatin, a G4 stabilizing ligand.[14] We chose these cell lines because they exhibited low to moderate expression of HOTAIR. Exposure to Pyridostatin (5 μM) for 24 h increased the expression of HOTAIR-N in all three cell lines (Figures 6A–6C). We postulated that G4 enhances the HOTAIR-N expression. Thus, we exposed the HOTAIR-intermediate H23 and HOTAIR-high T47D cells to TAP1 (0.5 μM & 2 μM), a ligand that can disrupt G4.[35] Indeed, exposure to TAP1 (2 μM) for 48 h reduced the expression of HOTAIR-N by 80% in both cell lines (Figures 6D and 6E). Moreover, G4 formation of HOTAIR-N G4-60 and G4-41 was disrupted by TAP1 in a dose-dependent manner (0–120 μM) as demonstrated by reduced peak intensities at 240 and 260 nM using CD assays (Figures 6F and 6G). It is noteworthy that Pyridostatin and TAP1 exerted opposite effects on the expression of HOTAIR-N in the same line, H23 cells in which the expression of HOTAIR-N was intermediate when compared with MDA-MB-231, A549, and T47D cells (Figures 6C and 6D). Consistent with TAP1's ability to disrupt G4 structure, TAP1 reduced the G4-60 harboring

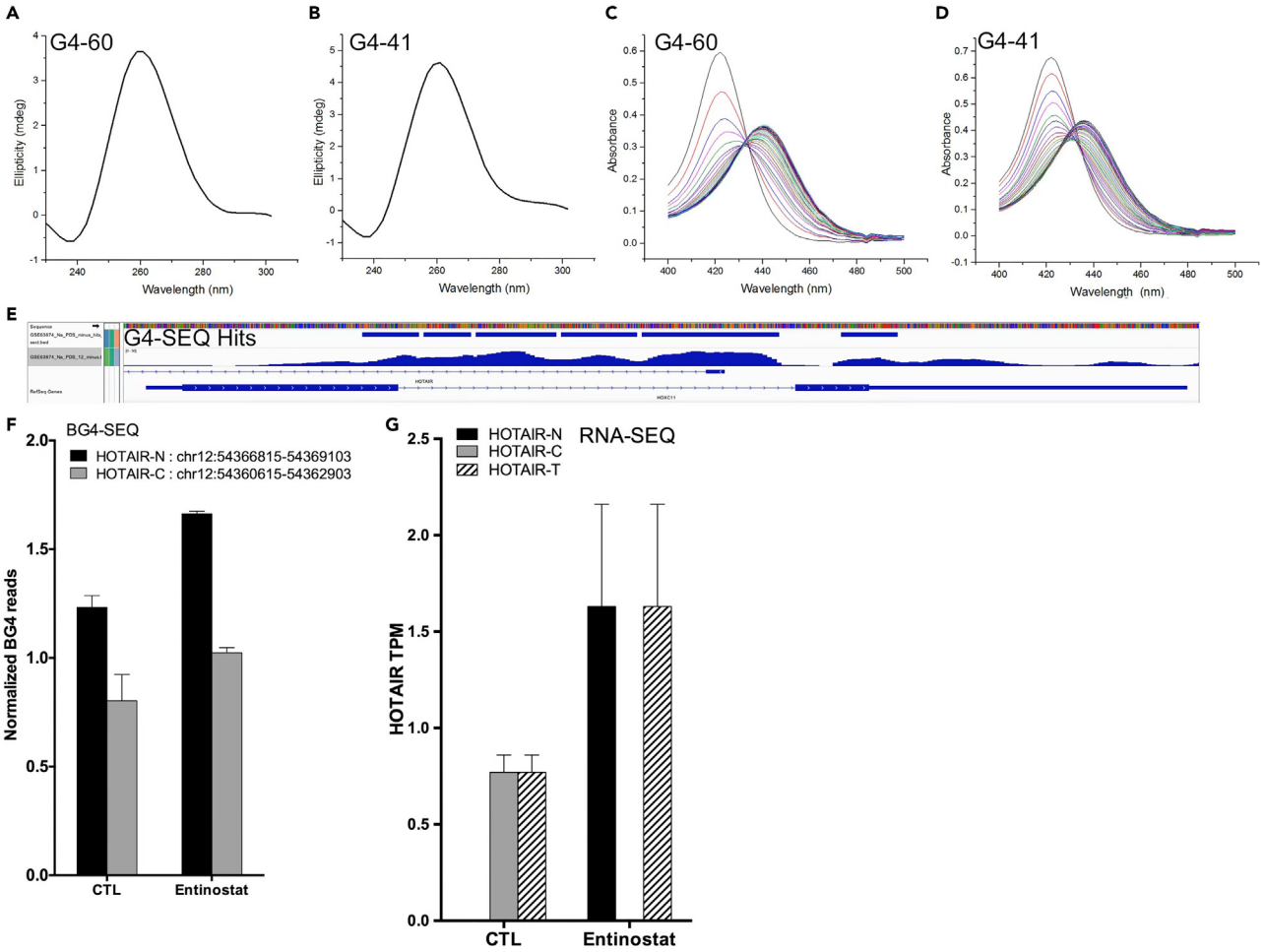


Figure 5. G4 motifs in the HOTAIR-N CpG island

(A) CD spectrum of the G4-60 oligo was assessed at a concentration of 10 μM in 10 mM Tris, 80 mM KCl buffer (pH 7.5).

(B) Similar to part A except that the CD spectra are assessed for G4-41.

(C) The UV-vis absorption spectra of TMPyP4 (5 μM) were measured upon titration with G4-60 (from 0 to 6.5 μM) in 10 mM Tris, 80 mM KCl buffer at pH 7.5.

(D) The UV-vis absorption spectra of TMPyP4 (5 μM) were measured upon titration with G4-41 (from 0 to 6.5 μM) in 10 mM Tris, 80 mM KCl buffer at pH 7.5.

(E) The G4 SEQ raw reads and peaks that were identified within the HOTAIR-N CpG island were visualized on IGV genome browser.

(F) Abundance of the BG4 bound HOTAIR-N CpG island and the corresponding region in the HOTAIR-C locus was quantified by extracting and analyzing genome wide BG4 ChIP sequencing data (BG4-SEQ) obtained from HACAT cells treated with Entinostat, an HDAC inhibitor.

(G) The amount of total HOTAIR (HOTAIR-T), HOTAIR-N, and HOTAIR-C was assessed by extracting and analyzing RNA-SEQ obtained from HACAT cells with or without exposure to an HDAC inhibitor Entinostat. Each transcript was quantified as transcript per million reads (TPM).

region bound by a G4-specific antibody BG4 by ~80% in H23 cells as demonstrated by BG4 ChIP assays (Figure 6H). To determine the direct effects of TAP1 on transcription from the HOTAIR-N G4 rich region, we transfected into T47D cells the HOTAIR-N-GLUC reporter that harbors the HOTAIR-N G4 motifs. Indeed, reporter activity of HOTAIR-N-GLUC was reduced by 72% after 48 h exposure to TAP1 (2 μM) (Figure 6I).

Regulation of HOTAIR-N expression by Bloom

G4 is unwound and resolved by DNA helicases, such as Bloom (BLM) that belongs to the RecQ family.[40] Interestingly, we noticed an inverse correlation between the expression of HOTAIR and BLM in the TCGA RNA-SEQ data of lung and breast tumors as determined by cBIOPORTAL (Table S19).[41] A similar trend was observed between HOXC11 and BLM. We also observed an inverse correlation between HOTAIR and BLM in Affymetrix gene expression array of the established human lung and breast cancer cell lines (GSE32474) (Table S20). It is noteworthy that BLM exhibited the highest expression in the HOTAIR-low MDA-MB-231 and lowest expression in the HOTAIR-high T47D cells in the array. The inverse correlation between HOTAIR-N, HOXC11, and BLM was further confirmed in our own RNA-SEQ of A549 and H23 cells (Figure 7A).

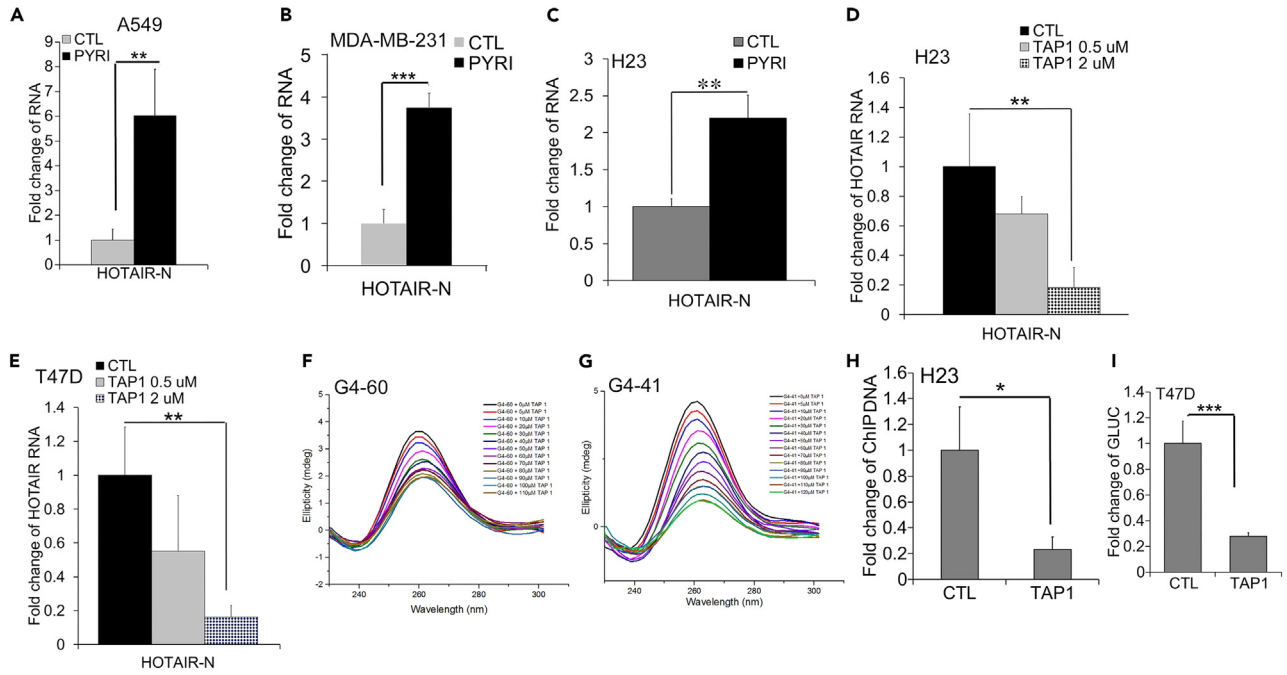


Figure 6. Regulation of the expression of HOTAIR-N by G4 ligands

(A) A549 cells were exposed to Pyridostatin (PYRI), a G4 stabilizing ligand (5 μM) for 24 h. Total RNA was extracted from the treated cells and assessed for the RNA levels of HOTAIR-N using real-time qPCR. A fold change was obtained by normalizing to the house keeping gene GAPDH and setting the values from the DMSO group to one (CTL).

(B) Similar to part A except that the RNA levels of HOTAIR-N were measured in MDA-MB-231 cells exposed to PYRI.

(C) Similar to part A except that the RNA levels of HOTAIR-N were measured in H23 cells exposed to PYRI.

(D) H23 cells were exposed to the G4 disruptive ligand TAP1 (0.5 & 2 μM) for 48 h. Total RNA was extracted and assessed for the RNA levels of HOTAIR-N using real-time qPCR. A fold change was obtained by normalizing to the house keeping gene GAPDH and setting the values from the DMSO group to one (CTL).

(E) Similar part D except that T47D cells were exposed to TAP1.

(F) CD spectra of G4-60 oligo assessed at a concentration of 10 μM in a buffer containing the indicated doses of TAP1 (0–110 μM) in 10 mM Tris, 80 mM KCl (pH 7.5).

(G) CD spectra of G4-41 oligo assessed at a concentration of 10 μM in a buffer containing the indicated doses of TAP1 (0–120 μM) in 10 mM Tris, 80 mM KCl (pH 7.5).

(H) G4 formation was assessed using ChIP assays with BG4, a G4-specific antibody in H23 cells treated with TAP1 (2 μM) for 48 h. The BG4 bound HOTAIR-N G4 dense region was quantified using PCR.

(I) The HOTAIR-N-GLUC and SEAP reporters were transfected into T47D cells followed by a 48 h exposure to TAP1 (2 μM). The ratios of HOTAIR-N-GLUC over SEAP were compared between the TAP1 treated and the control DMSO (CTL) groups. A fold change of the GLUC/SEAP ratios was obtained by setting the values from the CTL group to one. Mean and standard deviations were obtained from three independent experiments. *, **, and *** indicate a p value <0.05, 0.01, and 0.001, respectively.

To determine the role of BLM in regulation of HOTAIR-N expression we treated A549 and MDA-MB231 cells with ML216, an inhibitor of BLM that increases formation of G4.[17] Exposure to ML216 (25 μM) for 24 h substantially increased the RNA levels of HOTAIR-N and HOXC11 in the HOTAIR-low A549 and MDA-MB-231 cells (Figures 7B and 7C). Then we knocked down BLM using two distinct BLM-specific siRNAs (BLMsi) in A549 cells. Both BLMsi reduced the protein levels of BLM and increased the RNA levels of HOTAIR-N and HOXC11 in A549 cells (Figures 7D and 7E). We carried out RNA-SEQ on BLMsi-1 and CTLsi treated A549 cells. Knockdown of BLM significantly altered the expression of 4074 genes as illustrated by the heatmap (Figure 7F; Table S21). The panel was highlighted by the established G-quadruplex harboring genes (Table S21). For instance, the RNA levels of TERT were reduced by BLMsi to only 2% of that in the CTLsi group (Table S21). We then questioned whether BLMsi altered the expression of the G4 harboring genes globally. To assess enrichment of the G4 harboring genes altered by BLMsi we employed the pipe line that analyzed the G4 harboring genes in BLM deficient cells.[40] We extracted the genes harboring G4 motifs within 1000 bp either upstream or downstream of the transcription initiation site, which yielded 16914 genes (Table S22). Among them 4965 genes were significantly altered greater than 1.5-fold by BLMsi-1 (Figure 7G; Table S23). We analyzed the pathways and networks altered by BLMsi using Ingenuity Pathway. Because of the sheer large number of genes altered by BLMsi, we limited our scope to 500 most increased and 500 most decreased genes. The canonical pathways altered by BLMsi featured the pathways for cell cycle and DNA damage repair (Figure 7H; Table S24). Consistently, the networks altered by BLMsi were highlighted by Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair (Figure 7I; Table S25). GO analyses also revealed enrichment of DNA Replication,

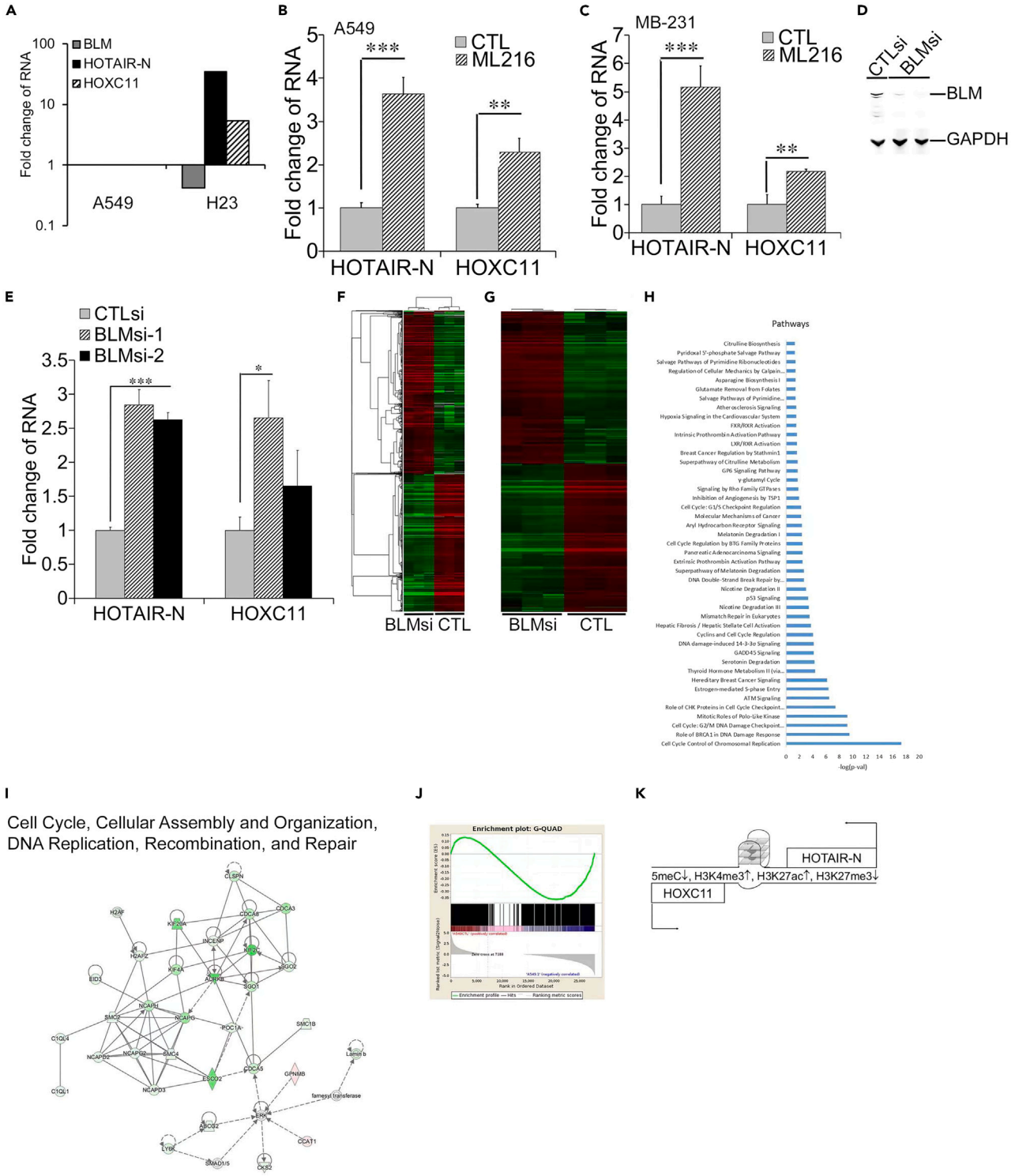


Figure 7. Increased expression of HOTAIR-N by inhibition of BLM

(A) The transcripts of HOTAIR-N, HOXC11, and BLM were quantified as transcript per million reads (TPM) using RNA-SEQ of A549 and H23 cells.

(B) A549 cells were exposed to ML216, an inhibitor of BLM at 25 μM for 24 h. Total RNA was extracted and assessed for the RNA levels of HOTAIR-N and HOXC11 using real-time qPCR. A fold change was obtained by normalizing to the house keeping gene GAPDH and setting the values from the DMSO group to one (CTL).

(C) Similar to part A except that MDA-MB-231 cells were exposed to ML216.

Figure 7. *Continued*

(D) Total cell lysates were extracted from A549 cells transfected with either a BLM-specific siRNA (BLMsi) or a control siRNA (CTLsi). The protein levels of BLM were assessed using immunoblots.

(E) The culture condition was similar to part B. Total RNA was extracted and the RNA levels of HOTAIR-N and HOXC11 were assessed using real-time qPCR. A fold change was obtained by normalizing to GAPDH and setting the values from the control siRNA group (CTLsi) to one.

(F) A549 cells were transfected with either CTLsi or BLM-Nsi. Total cell RNA were extracted and processed for RNA-SEQ. The differentially expressed gene set between CTLsi and BLMsi was established from RNA-SEQs of the triplicates of each group. The differentially expressed gene set was illustrated in a heatmap.

(G) The G4 harboring genes were identified in the gene set that were differentially expressed between CTLsi and BLMsi transfected A549 cells (as seen part F). The G4 harboring genes that were significantly altered by BLMsi were illustrated in a heatmap.

(H) The most significantly altered pathways by BLMsi were identified and illustrated using Ingenuity Pathway after filtering with a fold change >2 and an FDR <0.01.

(I) Ingenuity Pathway was employed to to identify the cancer related networks that were significantly altered by BLMsi in A549 cells through analysis of the RNA-SEQ data as seen in part F. The representative altered networks of Cell Cycle, DNA replication, Recombination, and Repair was illustrated using Ingenuity Pathway.

(J) GSEA was carried out to assess enrichment of G4 associated oncogene sets and tumor suppressor gene sets in the differentially expressed gene set between CTLsi and BLMsi transfected A549 cells (as seen in part F). A snap shot of the enrichment of the G4 gene set was presented. The complete presentation of the analysis was provided in Table S27.

(K) Epigenetic regulation of HOTAIR-N was illustrated in a working model. An open chromatin stage that mediates the expression of HOTAIR-N is achieved through reduced 5meC and H3K27me3, and increased H3K4me3, H3K27ac, and G4 formation. Mean and standard deviations were obtained from three independent experiments. ** and *** indicate a p value <0.01 and 0.001, respectively.

Recombination, and Repair functional terms (Table S26). In addition, GSEA analysis revealed that BLMsi altered the G4-regulated oncogene and tumor suppressor gene set as well as several epigenetic gene sets (Figure 7J; Table S27).

DISCUSSION

Investigation of upregulation of HOTAIR expression in cancer has mostly been focused on HOTAIR-C and the regulatory elements proximal to the HOTAIR-C transcription start site.[9,10] Herein, we demonstrate that HOTAIR-N is the driver isoform in lung and breast cancer cells (Figures 1 and 2). This conclusion is supported by our assessment of proliferation and viability, and by the gene expression program, signaling pathways, and networks that were altered by HOTAIR-Nsi (Figure 2). Moreover, the target genes of HOTAIR-N are highlighted with the genes categorized in the molecular mechanisms of cancer, such as CASP10, CDK6, JUN, E2F7, and TGFB1 (Table S10). These genes are crucial to cell cycle, apoptosis, and microenvironment of cancer cells.

HOTAIR-N regulates gene expression program likely via epigenetic mechanisms because the HOTAIR-N-regulated genes are enriched with target genes of 5meC, H3K27m3, and EZH2. Moreover, HOTAIR-N and HOXC11 form a sense-antisense gene pair and their expression is simultaneously elevated in tumor tissues and cancer cells from breast and lung despite their head-to-head configuration (Figure 1 & unpublished observations). These observations implicate potential reciprocal *cis* regulation within the HOTAIR-N-HOXC11 locus. Potential *cis* actions of HOTAIR-N are appealing because recent reports suggest HOTAIR regulates its neighbor HOXC genes in mice and HOXC11 itself promotes breast cancer via its interaction with steroid receptor coactivator-1.[42,43] The convergent HOTAIR-N-HOXC11 locus characterized herein presents an ideal platform to study dysregulation of sense-antisense gene pairs in cancer.

The HOTAIR-N CpG island is critical to the upregulation of HOTAIR-N expression in cancer cells because the reporter gene controlled by the HOTAIR-N CpG island is more active in the HOTAIR-high T47D cells than the HOTAIR-low MDA-MB-231 cells (Figure 3). The HOTAIR-N CpG island is also associated with a broad range of epigenetic markers of active transcription, namely low 5meC, and high H3K4me3, H3K27ac, and ATAC, which collectively indicate an open chromatin in the HOTAIR-N CpG island (Figure 3). In contrast, these markers in the HOTAIR-C corresponding region exhibit modest to minimal correlation with the expression of total HOTAIR (Figure 3). Moreover, inhibition of DNMT and HDAC can increase the expression of HOTAIR-N (Figure 4). These findings support a critical role of DNA and histone modifications in regulation of HOTAIR-N. HOTAIR promotes cancer via its interactions with various partners.[5] Thus, silencing the expression of HOTAIR appears to be an efficient approach to target HOTAIR in cancer. Our comparison of the HOTAIR-N and HOTAIR-C loci suggests that the epigenetic elements in the HOTAIR-N CpG island are the appealing targets to achieve this goal.

We identified G4 motifs in the HOTAIR-N CpG island (Figures 5, 6, and 7). We postulate that G4 formation mediates the expression of HOTAIR-N in cancer cells. Enhancing G4 formation has been extensively investigated as a therapeutic approach in cancer because G4 formation was associated with suppression of the expression of several oncogenes, such as MYC, KRAS, and KIT.[14] Development of G4 enhancing molecules for cancer therapy is largely based on the premise that accumulation of G4 in the oncogenes' promoters can silence the expression of oncogenes. However, our characterization of G4s in the HOTAIR-N CpG island challenges this oversimplified paradigm. Our results indicate that G4 formation is critical for upregulation of HOTAIR in cancer cells. This notion is based on our findings that enhancing G4 formation by the stabilizing ligands and inhibition of the DNA helicase BLM can increase the expression of HOTAIR-N, whereas disrupting G4 formation by TAP1 can decrease the expression of HOTAIR-N (Figures 6 and 7). Therefore, it is imperative to thoroughly understand how G4 upregulates and downregulates genes in a genomic context dependent manner. A recent integrative genome wide analysis of G4, epigenome, and transcriptome indicates that G4 is associated with active transcription and open chromatin globally.[13] We demonstrate such an association in the HOTAIR-N-HOXC11 locus (Figures 3, 4, 5, 6, and 7). Our results suggest that G4 regulates gene expression from the HOTAIR-N-HOXC11 locus epigenetically via direct interactions with DNMT and histone methyltransferase as recently reported.[44,45]

We propose that the genomic context embedding G4 may dictate whether a G4 upregulates or downregulates gene expression. G4s are commonly located in the promoters of the oncogenes that are suppressed by G4s.[14,46] In contrast, five G4 motifs in the HOTAIR-N gene are located in the first intron of HOTAIR-N (Figure S5). Another distinct feature of the HOTAIR-N G4s is that they are within a CpG island that HOTAIR-N and HOXC11 are transcribed in a head-to-head fashion (Figure 1A). This configuration resembles another sense-antisense gene pair Vimentin and VIM-AS1. The expression of Vimentin and VIM-AS1 are co-upregulated by R-loop, another non-B form single-stranded DNA structure.[47] In both sense-antisense gene pairs the transcribed CpG islands are enriched with G4 and R-loop motifs in the first intron (Figures S5 and S6). Moreover, G4 and R-loop motifs are either overlapping or proximal to each other (Figures S5 and S6). We observed increased expression of HOTAIR-N and HOXC11 by the topoisomerase I inhibitors camptothecin and topotecan in A549 cells, which provides further similarities between two loci (Figure S7).[47] Thus, G4 motifs might mediate gene expression from sense-antisense gene pairs via its interaction with epigenetic modifications and R-loop formation. The G4 and R-loop motifs in the HOTAIR-N CpG island might also facilitate gene translations that result in the NUP-98-HOXC11 gene fusion in leukemia because G4 and R-loop promote genome instability.[48]

G4s are resolved by DNA helicases in the cell.[49] Patients with deficiencies of these helicases, such as BLM and Warner, are cancer prone as a result of increased genome instability.[49] However, how G4 helicases regulate expression of G4 harboring genes in cancer remains unclear. We demonstrate an inverse correlation between BLM and HOTAIR in cancer cells and tumor tissues (Tables S19 and S20). Inhibition of BLM can increase the expression of HOTAIR-N and HOXC11, and regulate the cancer related gene expression programs and networks in cancer cells (Figure 7). Thus, we propose that deficiency of G4 helicases predisposes patients to cancer via upregulation of oncogene expression besides promotion of genome instability.

In summary, we identify HOTAIR-N as a dominant isoform in cancer with respect to its expression and functions. We report a critical role for G4 in upregulation of HOTAIR-N expression in cancer cells. Our results highlight a crosstalk among the epigenetic codes and implicate that G4 integrates epigenetic codes to upregulate gene expression (Figure 7K). We also provide the transcriptomes regulated by HOTAIR-N and BLM that are important resources for exploration of lncRNA, DNA helicase, and G4 in cancer cells. We propose a gene context dependent therapeutic enhancement or disruption of G4 in cancer rather than the current approach that aims solely at enhancement of G4.

Limitations of the study

Biological significance of the elevated expression of HOTAIR-N in cancer needs to be further explored and validated in animal models and clinical tissues of cancer. G-quadraplex needs to be precisely targeted to further understand its role in the activation of HOTAIR-N in cancer cells.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines
  - Cell culture
  - Reagents and plasmids
  - RNA extraction and quantitative RT-PCR
  - 5′RACE
  - Transfection
  - Cell proliferation and viability assay
  - Reporter assays
  - RNA-SEQ analysis
  - GRO-SEQ analysis
  - Functional analyses of transcriptomes
  - BS-SEQ analysis
  - ChIP-SEQ analysis
  - ATAC-SEQ analysis
  - *In vitro* G4 formation and ligand binding studies
  - Chromatin immunoprecipitation for G-quadruplex motifs
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108559.

ACKNOWLEDGMENTS

This work was supported by Washington State University Startup Fund awarded to B.S., Natural Science Foundation of Liaoning Province (2021-MS-168) to X.Q. This work was also supported in part by an NIH R01 (R01GM097571) to J.J., NIH COBRE grant (P20 GM121288) and a Tulane school of medicine faculty research pilot grant to Z.L. The Balasubramanian laboratory is supported by a program grant (C9681/A18618) and core funding (C14303/A17197) from Cancer Research UK.

We are also grateful to Dr. Weihang Chai for critical review and advice to this manuscript. We also appreciate Dr. Lucia Peixoto's guidance on bioinformatic analyses.

AUTHOR CONTRIBUTIONS

X.Q. carried out most of the experiments and wrote the manuscript. Z.L. carried out transcriptome and methalome analysis. J.J. carried out *in vitro* G-quadruplex analysis. J.S.A. carried out functional analysis of transcriptomes. E.S. carried out *in vitro* G-quadruplex analysis. K.P.N. carried out functional analysis of transcriptomes. F.F. carried out functional analysis of transcriptomes. S.B. contributed to disruption of G-quadruplex. B.S. conceived the study and revised the manuscript.

DECLARATION OF INTERESTS

S.B. is a founder, adviser, and shareholder of Cambridge Epigenetix Ltd.

Received: April 24, 2023
Revised: July 29, 2023
Accepted: November 20, 2023
Published: November 22, 2023

REFERENCES

1. Zhang, Y., Wang, C., Zou, X., Tian, X., Hu, J., and Zhang, C.Y. (2021). Simultaneous Enzyme-Free Detection of Multiple Long Noncoding RNAs in Cancer Cells at Single-Molecule/Particle Level. Nano Lett. *21*, 4193–4201.
2. Zhang, Y., Du, X.K., Liu, W.J., Liu, M., and Zhang, C.Y. (2022). Programmable Ligation-Transcription Circuit-Driven Cascade Amplification Machinery for Multiple Long Noncoding RNAs Detection in Lung Tissues. Anal. Chem. *94*, 10573–10578.
3. Zhao, N.N., Yu, X.D., Tian, X., Xu, Q., and Zhang, C.Y. (2023). Mix-and-Detection Assay with Multiple Cyclic Enzymatic Repairing Amplification for Rapid and Ultrasensitive Detection of Long Noncoding RNAs in Breast Tissues. Anal. Chem. *95*, 3082–3088.
4. Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.
5. Loewen, G., Jayawickramarajah, J., Zhuo, Y., and Shan, B. (2014). Functions of lncRNA HOTAIR in lung cancer. J. Hematol. Oncol. *7*, 90.
6. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.
7. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.
8. Bhan, A., Hussain, I., Ansari, K.I., Kasiri, S., Bashyal, A., and Mandal, S.S. (2013). Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. J. Mol. Biol. *425*, 3707–3722.
9. Özeş, A.R., Miller, D.F., Özeş, O.N., Fang, F., Liu, Y., Matei, D., Huang, T., and Nephew, K.P. (2016). NF-κB-HOTAIR axis links DNA damage response, chemoresistance and cellular senescence in ovarian cancer. Oncogene *35*, 5350–5361.
10. Pastori, C., Kapranov, P., Penas, C., Peschansky, V., Volmar, C.H., Sarkaria, J.N., Bregy, A., Komotar, R., St Laurent, G., Ayad, N.G., et al. (2015). The Bromodomain protein BRD4 controls HOTAIR, a long noncoding RNA essential for glioblastoma proliferation. Proc. Natl. Acad Sci. USA *112*, 8326–8331.
11. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.
12. Morgan, R., Hunter, K., and Pandha, H.S. (2022). Downstream of the HOX genes: Explaining conflicting tumour suppressor and oncogenic functions in cancer. Int. J. Cancer *150*, 1919–1932.
13. Hänsel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., et al. (2016). G-quadruplex structures mark human regulatory chromatin. Nat. Genet. *48*, 1267–1272.
14. Hänsel-Hertsch, R., Di Antonio, M., and Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat. Rev. Mol. Cell Biol. *18*, 279–284.
15. Ramirez, R.D., Sheridan, S., Girard, L., Sato, M., Kim, Y., Pollack, J., Peyton, M., Zou, Y., Kurie, J.M., Dimaio, J.M., et al. (2004). Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. Cancer Res. *64*, 9027–9034.
16. Sun, Z., Asmann, Y.W., Kalari, K.R., Bot, B., Eckel-Passow, J.E., Baker, T.R., Carr, J.M., Khrebtukova, I., Luo, S., Zhang, L., et al. (2011). Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. PLoS One *6*, e17490.
17. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323.
18. Strong, M.J., Baddoo, M., Nanbo, A., Xu, M., Puetter, A., and Lin, Z. (2014). Comprehensive high-throughput RNA sequencing analysis reveals contamination of multiple nasopharyngeal carcinoma cell lines with HeLa cell genomes. J. Virol. *88*, 10696–10704.
19. Li, M., Li, X., Zhuang, Y., Flemington, E.K., Lin, Z., and Shan, B. (2017). Induction of a novel isoform of the lncRNA HOTAIR in Claudin-low breast cancer cells attached to extracellular matrix. Mol. Oncol. *11*, 1698–1710.
20. Kent, W.J. (2002). BLAT–the BLAST-like alignment tool. Genome Res. *12*, 656–664.
21. Franco, H.L., Nagari, A., Malladi, V.S., Li, W., Xi, Y., Richardson, D., Allton, K.L., Tanaka, K., Li, J., Murakami, S., et al. (2018). Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. Genome Res. *28*, 159–170.
22. Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.
23. Fang, F., Cardenas, H., Huang, H., Jiang, G., Perkins, S.M., Zhang, C., Keer, H.N., Liu, Y., Nephew, K.P., and Matei, D. (2018). Genomic and Epigenomic Signatures in Ovarian Cancer Associated with Resensitization to Platinum Drugs. Cancer Res. *78*, 631–644.
24. Antoon, J.W., Lai, R., Struckhoff, A.P., Nitschke, A.M., Elliott, S., Martin, E.C., Rhodes, L.V., Yoon, N.S., Salvo, V.A., Shan, B., et al. (2012). Altered death receptor signaling promotes epithelial-to-mesenchymal

transition and acquired chemoresistance. Sci. Rep. *2*, 539.
25. Beacon, T.H., Delcuve, G.P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A.J., and Davie, J.R. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. Clin. Epigenetics *13*, 138.
26. Messier, T.L., Boyd, J.R., Gordon, J.A., Stein, J.L., Lian, J.B., and Stein, G.S. (2016). Oncofetal Epigenetic Bivalency in Breast Cancer Cells: H3K4 and H3K27 Tri-Methylation as a Biomarker for Phenotypic Plasticity. J. Cell. Physiol. *231*, 2474–2481.
27. Spangle, J.M., Dreijerink, K.M., Groner, A.C., Cheng, H., Ohlson, C.E., Reyes, J., Lin, C.Y., Bradner, J., Zhao, J.J., Roberts, T.M., et al. (2016). PI3K/AKT Signaling Regulates H3K4 Methylation in Breast Cancer. Cell Rep. *15*, 2692–2704.
28. Masalha, M., Ben-Dov, I.Z., Ram, O., Meningher, T., Jacob-Hirsch, J., Kassem, R., Sidi, Y., and Avni, D. (2021). H3K27Ac modification and gene expression in psoriasis. J. Dermatol. Sci. *103*, 93–100.
29. Wang, Y., Zhang, T., Kwiatkowski, N., Abraham, B.J., Lee, T.I., Xie, S., Yuzugullu, H., Von, T., Li, H., Lin, Z., et al. (2015). CDK7-dependent transcriptional addiction in triple-negative breast cancer. Cell *163*, 174–186.
30. Rhie, S.K., Hazelett, D.J., Coetzee, S.G., Yan, C., Noushmehr, H., and Coetzee, G.A. (2014). Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. BMC Genom. *15*, 331.
31. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.
32. Handoko, L., Kaczkowski, B., Hon, C.C., Lizio, M., Wakamori, M., Matsuda, T., Ito, T., Jeyamohan, P., Sato, Y., Sakamoto, K., et al. (2018). JQ1 affects BRD2-dependent and independent transcription regulation without disrupting H4-hyperacetylated chromatin states. Epigenetics *13*, 410–431.
33. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
34. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., and Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. Nat. Rev. Mol. Cell Biol. *21*, 459–474.
35. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. *34*, W676–W682.
36. Lam, E.Y., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013). G-quadruplex structures are stable and detectable in human genomic DNA. Nat. Commun. *4*, 1796.
37. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A., and Spada, G.P. (2010). A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. Org. Biomol. Chem. *8*, 2683–2692.
38. Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The Structure and Function of DNA G-Quadruplexes. Trends Chem. *2*, 123–136.
39. Pavlova, A.V., Kubareva, E.A., Monakhova, M.V., Zvereva, M.I., and Dolinnaya, N.G. (2021). Impact of G-Quadruplexes on the Regulation of Genome Integrity, DNA Damage and Repair. Biomolecules *11*, 1284.
40. Nguyen, G.H., Tang, W., Robles, A.I., Beyer, R.P., Gray, L.T., Welsh, J.A., Schetter, A.J., Kumamoto, K., Wang, X.W., Hickson, I.D., et al. (2014). Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. Proc. Natl. Acad. Sci. USA *111*, 9905–9910.
41. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.
42. Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. EMBO J. *36*, 981–994.
43. Amândio, A.R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. PLoS Genet. *12*, e1006232.
44. Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. Mol Cell *65*, 1056–1067.e1055.
45. Mao, S.Q., Ghanbarian, A.T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol. *25*, 951–957.
46. Linke, R., Limmer, M., Juranek, S.A., Heine, A., and Paeschke, K. (2021). The Relevance of G-Quadruplexes for DNA Repair. Int. J. Mol. Sci. *22*, 12599.
47. Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. USA *112*, 5785–5790.
48. Taketani, T., Taki, T., Shibuya, N., Kikuchi, A., Hanada, R., and Hayashi, Y. (2002). Novel NUP98-HOXC11 fusion gene resulted from a chromosomal break within exon 1 of HOXC11 in acute myeloid leukemia with t(11;12)(p15;q13). Cancer Res. *62*, 4571–4574.
49. Mendoza, O., Bourdoncle, A., Boulé, J.B., Brosh, R.M., Jr., and Mergny, J.L. (2016). G-quadruplexes and helicases. Nucleic Acids Res. *44*, 1989–2006.
50. Nguyen, H.T., Zhuang, Y., Sun, L., Kantrow, S.P., Kolls, J.K., You, Z., Zhuo, Y., and Shan, B. (2013). Src-mediated morphology transition of lung cancer cells in three-dimensional organotypic culture. Cancer Cell Int. *13*, 16.
51. Nguyen, H.T., Li, C., Lin, Z., Zhuang, Y., Flemington, E.K., Burow, M.E., Lin, Y.I., and Shan, B. (2012). The microRNA expression associated with morphogenesis of breast cancer cells in three-dimensional organotypic culture. Oncol. Rep. *28*, 117–126.
52. Nguyen, G.H., Dexheimer, T.S., Rosenthal, A.S., Chu, W.K., Singh, D.K., Mosedale, G., Bachrati, C.Z., Schultz, L., Sakurai, M., Savitsky, P., et al. (2013). A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. Chem Biol *20*, 55–62.
53. Waller, Z.A., Sewitz, S.A., Hsu, S.T., and Balasubramanian, S. (2009). A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. J. Am. Chem. Soc. *131*, 12628–12633.
54. Li, C., Nguyen, H.T., Zhuang, Y., Lin, Y., Flemington, E.K., Guo, W., Guenther, J., Burow, M.E., Morris, G.F., Sullivan, D., et al. (2011). Post-transcriptional up-regulation of miR-21 by type I collagen. Mol. Carcinog. *50*, 563–570.
55. Li, C., Nguyen, H.T., Zhuang, Y., Lin, Z., Flemington, E.K., Zhuo, Y., Kantrow, S.P., Morris, G.F., Sullivan, D.E., and Shan, B. (2012). Comparative profiling of miRNA expression of lung adenocarcinoma cells in two-dimensional and three-dimensional cultures. Gene *511*, 143–150.
56. Li, M., Li, X., Zhuang, Y., Wang, Y., Burow, M.E., Collins-Burow, B., Xue, M., Song, C., and Shan, B. (2016). Induction of HOXA9 expression in three-dimensional organotypic culture of the Claudin-low breast cancer cells. Oncotarget *7*, 51503–51514.
57. Shan, B., and Morris, G.F. (2005). Binding sequence-dependent regulation of the human proliferating cell nuclear antigen promoter by p53. Exp. Cell Res. *305*, 10–22.
58. Tannous, B.A. (2009). Gaussia luciferase reporter assay for monitoring biological processes in culture and in vivo. Nat. Protoc. *4*, 582–591.
59. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.
60. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589.
61. 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., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.
62. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.
63. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
64. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D., and Balasubramanian, S. (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. *13*, 551–564.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal antibody to Blooms Syndrome Protein	Abcam	Cat# ab2179; RRID: AB_2290411
Anti-DNA G-quadruplex structures monoclonal Antibody, clone BG4	Millipore Sigma	Cat# MABE917; RRID: AB_2750936
Chemicals, peptides, and recombinant proteins		
Inhibitor of Bloom ML216	Cayman Chemical	15186
Pyridostatin	Cayman Chemical	18013
Trichostatin A	Cayman Chemical	89730
5-Aza-2′-deoxycytidine	BioVision	Catalog #: 1754
TAP1	Department of Chemistry, University of Cambridge	NA
TMPyP4	Aldrich Sigma	613560
Critical commercial assays		
SMARTer™ RACE cDNA Amplification Kit	Takara Bio	Cat. Nos. 634858
Secrete-Pair™ Gaussia Luciferase Dual Luminescence Assay Kit	Genecopoeia	Cat. Nos. LF031
Deposited data		
RNA-SEQ Datasets	This study	GSE119510
RNA-SEQ Datasets	This study	GSE119511
RNA-SEQ Datasets	This study	GSE119512
RNA-SEQ Datasets	This study	GSE119513
RNA-SEQ Datasets	This study	GSE124934
Experimental models: Cell lines		
A549	ATCC	RRID: CVCL_UR31
H23	ATCC	RRID: CVCL_1547
MDA-MB-231	ATCC	RRID: CVCL_0062
T47D	ATCC	RRID: CVCL_0553
Oligonucleotides		
PCR primers and siRNAs used in this work, see Table S1	NA	NA
Software and algorithms		
Graphpad 6.0	Graphpad	NA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Xiaohan Qu (han_seal@163.com).

Materials availability

This study did not generate new unique materials.

Data and code availability

- RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- 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.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

The A549, H23, MDA-MB-231 and T47D cancer cell lines were purchased from ATCC and subcultured for our studies. All cell line studies were carried out within 15 passages. We did not independently authenticate our cell lines or test mycoplasma contamination.

Cell culture

The hTERT-CDK4-immortalized human bronchial epithelial cell line HBEC3 was kindly provided by Dr. Jerry Shay at University of Texas Southwestern Medical Center.[15] HBEC3 cells were cultured in RPMI1640.[15] Primary culture of normal human mammary epithelial cells (HMEC) was purchased from Lonza and cultured per the provider's instructions (Basel, Switzerland). Human lung cancer cell lines A549 and H23 were purchased from ATCC and cultured in RPMI1640 (Manassas, VA).[50] Human breast cancer cell lines MDA-MB-231 and T47D were purchased from ATCC and cultured in DMEM (Manassas, VA).[51] MCF-7-TNR, a chemoresistant variant of human breast cancer cell line MCF-7, was cultured DMEM.[24] These cell lines were chosen to investigate regulation of the expression of HOTAIR-N because they provided the context of minimal to robust expression of HOTAIR-N and benign to malignancy. A summary of the characters and salient results of these cell lines was provided in the supplemental information (Table S28).

Reagents and plasmids

Cayman Chemical (Ann Arbor, MI) provided ML216, an inhibitor of Bloom (BLM); Pyridostatin, a G4 stabilizing ligand; and Trichostatin A (TSA), a pan-HDAC inhibitor.[52] 5-Aza-2′-deoxycytidine (5-Aza-2dC), a DNA methyltransferease (DNMT) inhibitor, was purchased from BioVision (Milpitas, CA). A BLM-specific antibody (ab2179) was purchased from Abcam (Cambridge, MA). TMPyP4, a G4 interacting ligand was obtained from Aldrich Sigma (St. Louis, MO). TAP1, a G4 disrupting ligand, was provided by Dr. Shankar Balasubramanian.[53] Sequences of the primers and siRNAs were listed in Table S1.

RNA extraction and quantitative RT-PCR

Total cell RNA was extracted using Trizol (Invitrogen, Carlsbad CA) per the provider's instructions.[54] Reverse transcription was carried out using High-Capacity cDNA Reverse (Transcription Kit Invitrogen, Carlsbad CA). Real-time PCR was carried out using Applied Biosystems TaqMan Universal PCR Master Mix. The RNA levels of each gene were measured using qRT-PCR on StepOne Plus Thermal Cycler (Invitrogen, Carlsbad CA).[55] A fold change of each transcript was obtained by normalizing to GAPDH and setting the values from the control group to one.

5′RACE

DNA was extracted from T47D cells. 5′RACE was carried out to map the transcription initiation site of HOTAIR-N using SMARTer RACE cDNA Amplification Kit per the provider's instructions (Takara Bio, Mountainview CA). DNA was extracted from the resultant clones and sequenced by Eton Bioscience (San Diego, CA). The sequencing results were blasted against hg19; Genome Reference Consortium GRCH37 using NCBI's Oligonucleotide Blast engine. The blast results were visualized in Integrative Genomics Viewer (IGV) genome browser.

Transfection

The HOTAIR-N-specific siRNA was designed using IDT's siRNA designing tool and purchased from IDT (Coralville, IA). A second human HOTAIR-specific siRNA that targets all isoforms of HOTAIR (ID: SASI_Hs02_00380445) and the control siRNA were purchased from Sigma (St. Louis, MO). The human BLM-specific siRNA (BLMsi) and the control siRNAs (CTLsi) were purchased from GenePharma (Shanghai China). The indicated cancer cells were seeded into 6-well culture dishes at a density of $2.5x10^5$ cells/well. All the siRNAs were transfected at 60 nM into the indicated cell lines using RNAiMAX per the reverse transfection protocol (Invitrogen, Carlsbad CA).[56] RNA were extracted at 48 h after transfection for further analysis using Trizol agent.

Cell proliferation and viability assay

Cell Proliferation was assessed when T47D cells were transfected with either HOTAIR-NsiRNA or CTLsiRNA. After trypsinization live cells were counted using a hematocytometer coupled with trypan blue exclusion at 72 h after seeding. A fold change in cell count was determined by calculating the ratios of the cell number at 48 h after seeding over the number of cells seeded. Cell viability was assessed using MTT Assays per the provider's instructions (ThermoFisher, Waltham, MA).[57]

Reporter assays

GeneCopoeia provided the control reporter SEAP and the GLUC reporter construct in which expression of the GLUC reporter gene is controlled by the HOTAIR-N CpG island region (HOTAIR-N-GLUC).[58] The HOTAIR-N-GLUC and SEAP reporters were co-transfected into T47D and MDA-MB-231 cells. Briefly, the cells were split and seeded into a 24-well plate at a density of 5 x 10^4cells/well. On the next day, 50 ng of each reporter construct was co-transfected into the cells using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). The treated cells were harvested and assayed for reporter activity according to the supplier's instructions. The reporter activities were measured at 48 h after transfection. The ratios of GLUC/SEAP were compared between the transfected MDA-MB-231 and T47D cells.

RNA-SEQ analysis

Raw RNA-sequencing (RNA-SEQ) reads from invasive breast carcinoma (BRCA) and lung adenocarcinoma (LUAD) samples, and their paired normal samples were generated by The Cancer Genome Atlas (TCGA) project and obtained from the National Cancer Institute Genomic Data Commons that were reported in the previous publications.[59] The datasets were then analyzed using the RSEM algorithm for the quantification of isoforms of HOTAIR transcripts.[17,18] We stratified all samples into two groups based on whether HOTAIR-N was increased in tumor over their paired non-tumor tissues. For overall survival analysis we used cBioportal platform to carry out Kaplan-Meier Estimate and log -rank test (cBioportal.org).Raw RNA-SEQ reads from immortalized keratinocytes HACAT, MDA-MB-231, and MDA-MB-468 cells were downloaded from NCBI's Sequence Read Archive (GSE96867, SRA# SRP068243).[21] The datasets were then analyzed using STAR, RSEM, and EbSeq softwares.[17,18]

In house RNA-SEQ data were generated from A549, H23, T47D, and MCF-7-TNR cells at Washington State University Spokane Genomic Core. cDNA libraries were prepared from ribosomal RNA-depleted (ribodepleted) RNAs using the Illumina Truseq Stranded TotalRNASample Prep kit (RS-122-2101). The ribodepleted RNA cDNA library was subjected to 100-base paired end strand-specific sequencing on an Illumina HiSeq 2500。 Roughly 40 million paired end stranded 100 bp reads were generated from each RNA sample. The RNA-SEQ data can be accessed at NCBI GEO (GSE119510, GSE119511, GSE119512, GSE119513, GSE124934). Raw sequence data were aligned to a reference genome containing a human genome (hg19; Genome Reference Consortium GRCH37). The alignments were performed using Spliced Transcripts Alignment to a Reference (STAR) aligner version 2.3.0 and were subjected to visual inspection using the Integrative Genomics Viewer (IGV) genome browser. Transcript data from STAR were subsequently analyzed using RSEM version 1.2.09 for quantification of human gene expression. A transcript was defined as differentially expressed between two selected groups when its difference was greater than 2-fold and its false discovery rate (FDR) was smaller than 0.01.

GRO-SEQ analysis

GRO-SEQ data generated from two basal-like human breast cancer cell lines, MDA-MB-468 and MDA-MB-231 cells (GSE96867), were downloaded from NCBI's SRA.[21] GRO-SEQ reads were aligned to HG19 with BWA 0.7.17-r1188 using default settings. HOMER's makeTagDirectory was used for each GRO-SEQ file.[60] HOMER's makeUCSCfile was used on each GRO-SEQ file's Tag Directories twice, once with the "-stand +" option and once with "-strand -" options to create strand specific bedGraph files. BEDTOOLS v2.27.1 multicov and SAMTOOLS 1.8 view utilities were used to quantify the number of uniquely mapped reads in the HOTAIR-N CpG island and the total number of uniquely mapped reads across the genome, respectively.

Functional analyses of transcriptomes

Gene Set Enrichment Analysis (GSEA) was carried out to identify the gene sets regulated by the indicated treatments.[61] The RNA-SEQ data were quatified and normalized as in our RNA-SEQ analysis and used as input into our GSEA analysis. C2 curated gene sets were chosen for GSEA in the GSEA's MSigDB collection. We used a cutoff FDR value of 25% as recommended by GSEA's user guide.

BS-SEQ analysis

Raw whole-genome bisulfite sequencing (BS-SEQ) reads from T47D and MDA-MB-231 cells were obtained from the NCBI Sequence Read Archive (SRA# SRP005601) and analyzed with the Bismark Bisulfite Mapper (ver 0.16.1) using the default option.[62] The human genome hg19; Genome Reference Consortium GRCH37 was used as the reference genome. The percentage methylation within the HOTAIR CpG islands was compared between T47D and MDA-MB231 cells.

ChIP-SEQ analysis

Raw sequencing reads of ChIP-SEQ (H3K4me3, H3K27ac, BG4) generated from A549, H23, MDA-MB-231, and T47D cells were downloaded from the NCBI Sequence Read Archive and were aligned with the BWA aligner against the human genome assembly hg19.[63] Reads mapped to the HOTAIR-N and HOTAIR-C promoter regions were extracted based on their corresponding chromosome coordinates on the hg19 reference genome and the number of ChIP reads were further normalized by the total number of mapped ChIP reads and represented as "reads per million mapped ChIP reads".

ATAC-SEQ analysis

Raw ATAC sequencing reads of MDA-MB-231 and T47D cells were downloaded from NCBI SRA (SRP062544 and SRP078834) and aligned with the BWA aligner against the human genome assembly hg19. Reads mapped to the HOTAIR-N and HOTAIR-C promoter regions were extracted based on their corresponding chromosome coordinates on the hg19 reference genome and the number of ATAC reads were further normalized by the total number of mapped ATAC reads and represented as "reads per million mapped ATAC reads".

In vitro G4 formation and ligand binding studies

Oligonucleotides (oligos) G4-60 and G4-41 were synthesized by the Keck Foundation Biotechnology Resource Laboratory at Yale University using standard automated solid-phase synthesis. The oligos were desalted and subsequently purified using a Varian Prostar HPLC system. In order to facilitate G4 formation for circular dichroism spectroscopy (CD), oligos G4-60 or G4-41 were incubated in a micro-centrifuge tube

containing 10 mM Tris, 80 mM KCl, buffer (pH 7.5) at 90°C for 10 min followed by cooling to room temperature. The CD spectra of the G4s were obtained on an Olis RSM 1000 CD using a 1 mm path length circular quartz cuvette (Hellma) at room temperature. For the titration with TAP1: To 10 μM G4 oligos (10 mM Tris, 80 mM KCl, pH 7.5, buffer) was added increasing concentration of TAP1 until no further changes in the CD spectrum was observed.

UV–vis absorption studies were carried out on a Hewlett-Packard 8452A Diode Array Spectrophotometer using a 1400 μL, 10 mm path length, quartz cuvette (Nova Biotech), at room temperature. The G4 interacting ligand, TMPyP4, binding to G4s was investigated using UV-vis. Specifically, a solution of TMPyP4 (5 μM in 10 mM Tris, 80 mM KCl, pH 7.5 buffer) was titrated with a stock solution of 40 μM of G4 oligos and incubated for 1 min between readings until the total concentration of G4 oligos added was 6.5 μM.

Chromatin immunoprecipitation for G-quadruplex motifs

ChIP assays were employed to assess G4 formation in the HOTAIR-N G4 dense region. BG4, a G4-specific antibody was purchased from Millipore and used in ChIP assays to immunoprecipitated G4 motifs as.[64] Briefly, Crosslinked and sheared chromatin were prepared from $1x10^7$ cells. The samples were then immunoprecipitated with 1 μg of the BG4 antibody specific for G4, or a negative control antibody. The immunoprecipitated DNA was then recovered using a standard ethanol precipitation procedure. The precipitated DNA was used for real-time PCR to quantify enrichment of the HOTAIR-N G4 harboring motif. The sequences of the primers were provided in Table S1.

QUANTIFICATION AND STATISTICAL ANALYSIS

When presented, means and standard deviations were obtained from at least three independent experiments. A *p* value between any two compared groups was determined using unpaired two-tailed Student's *t* test (GraphPad Prism, Version 6).