Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis

Rajnish A. Gupta[1], Nilay Shah[4], Kevin C. Wang[1], Jeewon Kim[2], Hugo M. Horlings[6], David J. Wong[1], Miao-Chih Tsai[1], Tiffany Hung[1], Pedram Argani[5], John L. Rinn[7], Yulei Wang[8], Pius Brzoska[8], Benjamin Kong[8], Rui Li[3], Robert B. West[3], Marc J. van de Vijver[6], Saraswati Sukumar[4] & Howard Y. Chang[1]

Large intervening non-coding RNAs (lincRNAs) are pervasively transcribed in the genome[1–3] yet their potential involvement in human disease is not well understood[4,5]. Recent studies of dosage compensation, imprinting, and homeotic gene expression suggest that individual lincRNAs can function as the interface between DNA and specific chromatin remodelling activities[6–8]. Here we show that lincRNAs in the *HOX* loci become systematically dysregulated during breast cancer progression. The lincRNA termed *HOTAIR* is increased in expression in primary breast tumours and metastases, and *HOTAIR* expression level in primary tumours is a powerful predictor of eventual metastasis and death. Enforced expression of *HOTAIR* in epithelial cancer cells induced genome-wide re-targeting of Polycomb repressive complex 2 (PRC2) to an occupancy pattern more resembling embryonic fibroblasts, leading to altered histone H3 lysine 27 methylation, gene expression, and increased cancer invasiveness and metastasis in a manner dependent on PRC2. Conversely, loss of *HOTAIR* can inhibit cancer invasiveness, particularly in cells that possess excessive PRC2 activity. These findings indicate that lincRNAs have active roles in modulating the cancer epigenome and may be important targets for cancer diagnosis and therapy.

We hybridized RNA derived from normal human breast epithelia, primary breast carcinomas, and distant metastases to ultra-dense *HOX* tiling arrays[7] (Fig. 1a, b). We found that 233 transcribed regions in the *HOX* loci, comprising 170 non-coding RNAs (ncRNAs) and 63 *HOX* exons, were differentially expressed (Fig. 1a). Unsupervised hierarchical clustering showed systematic variation in the expression of *HOX* lincRNAs among normal breast epithelia, primary tumour, and metastases. *HOXA5*, a known breast tumour suppressor[9], along with dozens of *HOX* lincRNAs, are expressed in normal breast but with reduced expression in all cancer samples (Supplementary Fig. 1). A set of *HOX* lincRNAs and messenger RNAs, including the known oncogene *HOXB7* (ref. 10), is frequently expressed in primary tumours but not in metastases (Supplementary Fig. 1). A distinct set of *HOX* lincRNAs is sometimes overexpressed in primary tumours, and very frequently overexpressed in metastases (Fig. 1b). Notably, one such metastasis-associated lincRNA is *HOTAIR* (Fig. 1b), which has a unique association with patient prognosis (Supplementary Figs 1, 2 and Supplementary Table 1). *HOTAIR* is a lincRNA in the mammalian *HOXC* locus that binds to and targets the PRC2 complex to the *HOXD* locus, located on a different chromosome[7]. PRC2 is a histone H3 lysine 27 (H3K27) methylase involved in developmental gene silencing and cancer progression[11,12]. We proposed that altered *HOTAIR* expression may be involved in human cancer by promoting genomic relocalization of the Polycomb complex and H3K27 trimethylation.

Quantitative PCR showed that *HOTAIR* is overexpressed from hundreds to nearly two-thousand-fold in breast cancer metastases, and the *HOTAIR* expression level is sometimes high but heterogeneous among primary tumours (Fig. 1c). We next measured the *HOTAIR* level in an independent panel of 132 primary breast tumours (stage I and II) with extensive clinical follow-up[13]. Indeed, nearly one-third of primary breast tumours overexpress *HOTAIR* by more than 125-fold compared to normal breast epithelia, the minimum level of *HOTAIR* overexpression observed in bona fide metastases (Fig. 1d), and a high *HOTAIR* level is a significant predictor of subsequent metastasis and death ($P = 0.0004$ and $P = 0.005$ for metastasis and death, respectively, Fig. 1e, f). Multivariate analysis showed that prognostic stratification of metastasis and death by *HOTAIR* is independent of known clinical risk factors such as tumour size, stage and hormone receptor status (Supplementary Table 2).

We next examined the effects of manipulating *HOTAIR* level in several breast cancer cell lines. *HOTAIR* levels in cell lines are significantly lower than those seen in primary or metastatic breast tumours (Supplementary Figs 3 and 4). Retroviral transduction allowed stable overexpression of *HOTAIR* of several-hundred-fold compared to vector-transduced cells, which are comparable to levels observed in patients (Supplementary Fig. 4). *HOTAIR* overexpression promoted colony growth in soft agar (Supplementary Fig. 5). In addition, enforced expression of *HOTAIR* in four different breast cancer cell lines increased cancer cell invasion through Matrigel, a basement-membrane like extracellular matrix, (Fig. 2a). Conversely, depletion of *HOTAIR* by small interfering RNAs (siRNAs) in MCF7, a cell line that expresses endogenous *HOTAIR*, decreased its matrix invasiveness (Fig. 2b and Supplementary Fig. 6). To probe the effects of *HOTAIR* on cancer cell dynamics *in vivo*, we labelled control and *HOTAIR*-expressing cells with firefly luciferase, enabling *in vivo* bioluminescence imaging. When MDA-MB-231 cells expressing vector or *HOTAIR* were orthotopically grafted into mammary fat pads, serial imaging showed that *HOTAIR* expression modestly increased the rate of primary tumour growth (Fig. 2c, left). Notably, in the same animals, we observed significantly increased foci of luciferase signal in the lung fields of mice bearing $HOTAIR^+$ primary tumours (Fig. 2c, right), which suggests that *HOTAIR* promotes lung metastasis.

To quantify further metastatic potential *in vivo*, we performed tail vein xenografts and compared the rates of lung colonization. Vector expression in the non-metastatic cell line SK-BR3 never showed lung colonization after tail vein xenograft (0 out of 15 mice), but *HOTAIR*

[1]Howard Hughes Medical Institute and Program in Epithelial Biology, [2]Stanford Cancer Center and Transgenic Mouse Research Center, [3]Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA. [4]Sidney Kimmel Comprehensive Cancer Center, [5]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. [6]Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands. [7]The Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [8]Applied Biosystems, Foster City, California 94404, USA.

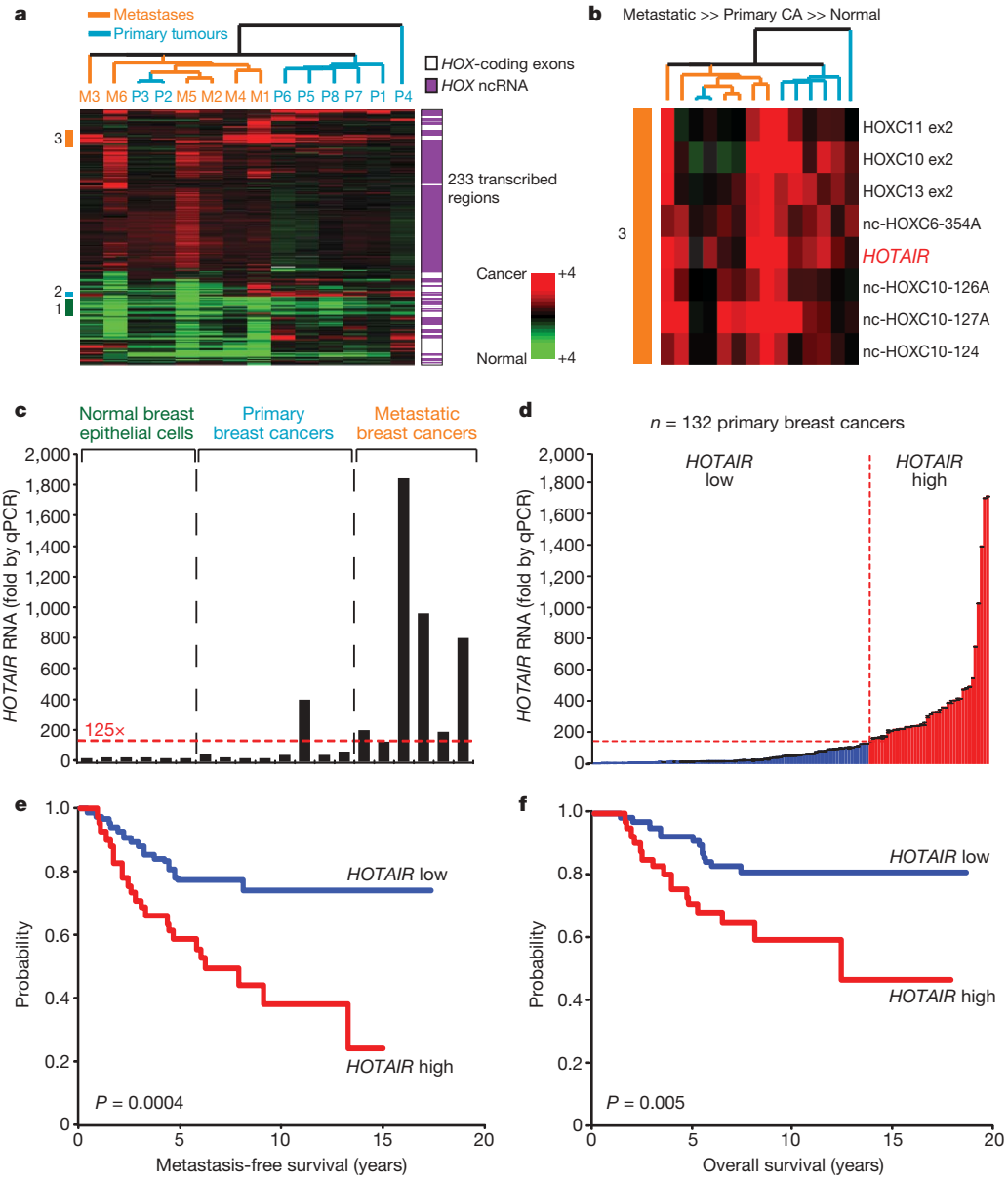

Figure 1 | *HOX* lincRNAs are systematically dysregulated in breast carcinoma and have prognostic value for metastasis and survival. **a**, Heat map representing unsupervised hierarchical clustering of expression values of a panel of primary and metastatic breast cancers relative to normal breast epithelial cells (pooled from five breast organoids). An ultra-high-density *HOX* tiling array[7] was interrogated with either normal breast organoid RNA (Cy3 channel) or RNA derived from primary or metastatic breast tumours (Cy5 channel). Each column represents the indicated clinical sample. Each row indicates a transcribed region, either a *HOX*-coding exon or *HOX* ncRNA. Expression values are depicted as the ratio relative to pooled normal and represented as a red–green colour scale. **b**, Higher resolution of subset 3 identifying transcripts that show higher relative expression in metastatic as compared to primary tumours and normal epithelia. *HOTAIR* is one such transcript ($P = 0.03$, Student's *t*-test). **c**, qRT–PCR validation of the expression tiling array results measuring *HOTAIR* abundance in a panel of normal breast epithelial-enriched organoids, primary breast tumours, and metastatic breast tumours. Metastatic tumours had a minimum of 125-fold higher levels of *HOTAIR* than normal breast epithelia. Error bars represent s.d. ($n = 3$). **d**, qRT–PCR analysis of *HOTAIR* in 132 primary breast tumours (stage I or II). Approximately one-third of primary breast tumours had >125-fold overexpression of *HOTAIR* compared to normal tissue (*HOTAIR* high, indicated in red), whereas roughly two-thirds of tumours did not (*HOTAIR* low, indicated in blue). Error bars represent s.d. ($n = 3$). **e**, **f**, Kaplan–Meier curves for metastasis-free survival (**e**) or overall survival (**f**) of the same 132 primary breast tumours measured in **d**.

expression allowed SK-BR3 cells to colonize the lung in 80% of animals (12 out of 15 mice, Fig. 2c). SK-BR3 cells apparently lack further genetic elements required to persist in the lung, because *HOTAIR*-transduced SK-BR3 cells in the lung disappeared after approximately 1 week. In contrast, *HOTAIR* expression in MDA-MB-231 cells resulted in approximately eight- to ten- fold more cells to engraft the lung after tail vein xenograft (Fig. 2d). These differences persisted until the end of the experiment, resulting in tenfold more lung metastases as verified by histology ($P = 0.00005$, Fig. 2e). The tumours retained *HOTAIR* expression for the length of the experiment (Supplementary Fig. 7).

We next tested whether *HOTAIR* overexpression affected the pattern of PRC2 occupancy. We mapped PRC2 occupancy genome-wide by chromatin immunoprecipitation followed by hybridization to tiling microarrays interrogating all human promoters (ChIP-chip, Fig. 3). Compared to vector-expressing cells, *HOTAIR* overexpression induced localization of H3K27me3 and PRC2 subunits SUZ12 and EZH2 on 854 new genes while concomitantly losing PRC2 occupancy and H3K27me3 on 37 genes (Fig. 3a). A significant fraction of these 854 genes also showed consequent changes in gene expression after *HOTAIR* overexpression (39% observed versus 7% expected by chance alone, $P = 2.5 \times 10^{-209}$, hypergeometric distribution). Most

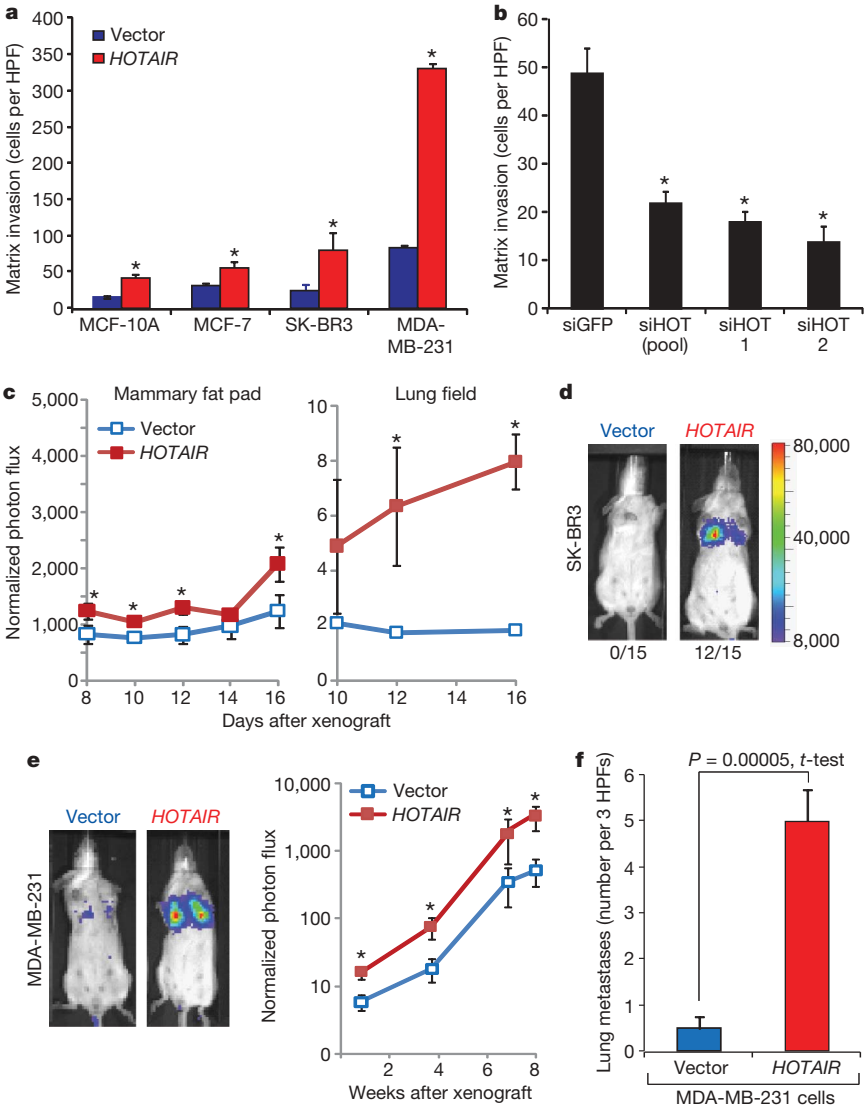

Figure 2 | *HOTAIR* promotes invasion of breast carcinoma cells. **a**, Relative fold increase in matrix invasion in four breast carcinoma cell lines after enforced *HOTAIR* expression. Mean ± s.d. are shown ($n = 3$). HPF, high power field. **b**, Matrix invasion in the MCF-7 breast carcinoma cell line transfected with individual or pooled siRNAs targeting *HOTAIR* (error bars = s.d., $n = 3$). siGFP denotes siRNA targeting green fluorescent protein (GFP) as a control. **c**, *HOTAIR* expression in MDA-MB-231 cells enhances orthotopic growth in mammary fat pads and metastasis to lung ($n = 5$ per arm; error bars are s.e.m.). **d**, *HOTAIR* promotes transient lung colonization of SK-BR3 after tail vein xenograft ($n = 15$ per arm). **e**, *HOTAIR* promotes lung colonization of MDA-MB-231 cells after tail vein xenograft. **f**, Histological analysis of the number of lung metastasis in vector- or *HOTAIR*-expressing MDA-MB-231 cells 8–9 weeks after tail vein xenograft ($n = 18$ per arm, error bars are s.e.m.). Similar results were obtained using luciferase-marked cells. *$P < 0.05$ between control cells and cells manipulated for *HOTAIR*.

PRC2 occupancy sites on promoters genome-wide showed little change (data not shown), and *HOTAIR* overexpression did not change the levels of PRC2 subunits (Fig. 4a, lane 1 versus lane 4). Several genes with *HOTAIR*-induced PRC2 occupancy are implicated in inhibiting breast cancer progression, including transcription factors *HOXD10* (ref. 14) and *PGR* encoding progesterone receptor (a classic favourable prognostic factor); cell adhesion molecules of the protocadherin (*PCDH*) gene family[15] and *JAM2* (ref. 16); and *EPHA1* (refs 17, 18), encoding an ephrin receptor involved in tumour angiogenesis. Gene Ontology[19] analysis suggested that most of the 854 genes are involved in pathways related to cell–cell signalling and development (Supplementary Fig. 8). *HOTAIR*-induced PRC2 occupancy tended to spread over promoters, and to a lesser extent, gene bodies (Fig. 3b). *HOTAIR* may also induce PRC2 localization to other intergenic regions not present on our tiling arrays. ChIP followed by quantitative PCR confirmed that *HOTAIR* substantially increased PRC2 occupancy and H3K27me3 of all target genes examined (Supplementary Fig. 9). Notably, like *HOTAIR* itself, the 854 *HOTAIR*–PRC2 target genes are coordinately downregulated in aggressive breast tumours that tend to cause death ($P < 0.0003$, Supplementary Fig. 10).

We next compared the 854 genes with *HOTAIR*-induced PRC2 occupancy in MDA-MB-231 cells with a compendium of published PRC2 occupancy profiles in diverse cell types (Fig. 3c). PRC2 occupancy patterns from different cancer, fibroblastic and embryonic stem cell lines were annotated from existing databases (Supplementary Table 3). Using a pattern-matching algorithm[20], we found that the *HOTAIR*-induced PRC2 occupancy pattern in breast cancer cells most closely resembled the endogenous PRC2 occupancy pattern in embryonic and neonatal fibroblasts, especially fibroblasts derived from posterior and distal anatomic sites where endogenous *HOTAIR* is expressed[7] ($P < 10^{-50}$ for each comparison, false discovery rate (FDR) $\ll 0.05$, Fig. 3c). These 854 genes are also significantly enriched for genes in primary fibroblasts that are bound by PRC2 in a *HOTAIR*-dependent manner (32% overlap observed versus 9.9% expected by chance alone, $P = 8.5 \times 10^{-93}$, hypergeometric distribution, M.-C.T., unpublished data). These results indicate that increased *HOTAIR* expression in breast cancer cells seems to reprogram the Polycomb

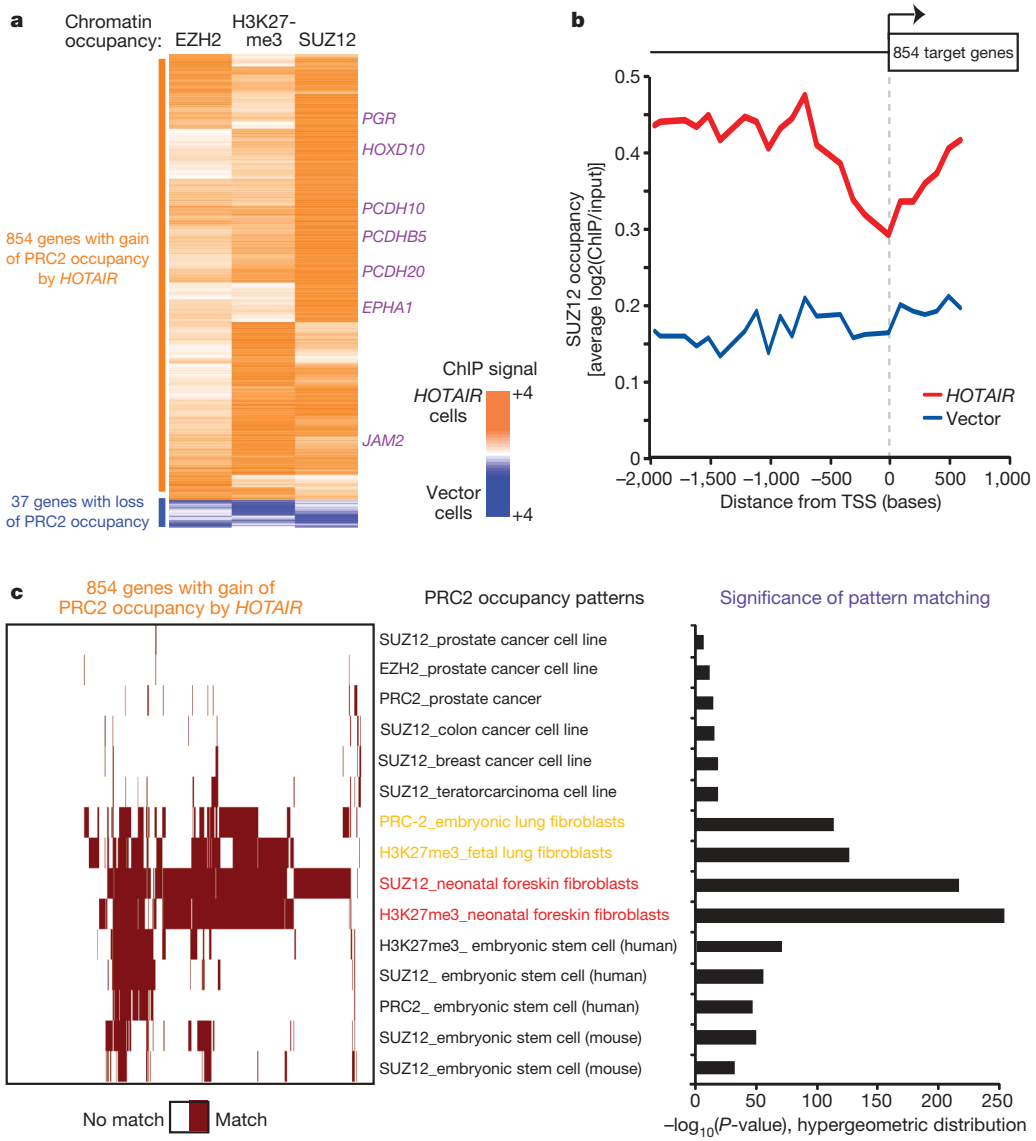

Figure 3 | *HOTAIR* promotes selective re-targeting of PRC2 and H3K27me3 genome-wide. a, Heat map representing genes with a significant relative change in chromatin occupancy of EZH2, SUZ12 and H3K27 after *HOTAIR* expression. MDA-MB-231 vector or *HOTAIR* cells were subjected to ChIP using anti-EZH2, -H3K27me3 and -SUZ12 antibodies followed by interrogation on a genome-wide promoter array. Values are depicted as the relative ratio of *HOTAIR* to vector cells and represented as an orange–blue scale. **b**, Average SUZ12 occupancy of >800 PRC2 target genes in *HOTAIR* or vector-expressing cells across the length of gene promoter and gene body. All target genes are aligned by their transcriptional start sites (TSS). **c**, Module map[20] of the 854 genes with a gain in PRC2 occupancy after *HOTAR* overexpression. Left, heat map of genes (column) showing a gain in PRC2 occupancy after *HOTAR* expression in breast carcinoma cells (see **a**) compared with PRC2 occupancy patterns from the indicated cell or tissue type (rows). Binary scale is brown (match) or white (no match). Right, quantification of significance of pattern matching between gene sets.

binding profile of a breast epithelial cell to that of an embryonic fibroblast.

Finally, we addressed whether the ability of *HOTAIR* to induce breast cancer invasiveness required an intact PRC2 complex. We transduced vector- or *HOTAIR*-expressing MDA-MB-231 cells with short hairpin RNAs (shRNAs) targeting PRC2 subunits EZH2 or SUZ12. Immunoblot analyses confirmed efficient depletion of the targeted proteins (Fig. 4a). Depletion of either SUZ12 or EZH2 had little effect on the invasiveness of control cells, but completely reversed the ability of *HOTAIR* to promote matrix invasion (Fig. 4b). Depletion of EZH2 also inhibited *HOTAIR*-driven lung colonization after tail vein xenograft by approximately 50% ($P < 0.05$). These results indicate that PRC2 is specifically required for *HOTAIR* to promote cellular invasiveness. Global gene expression analysis showed hundreds of genes that were induced or repressed as a consequence of *HOTAIR* overexpression (Fig. 4c, left). Importantly, concomitant depletion of PRC2 in large part reversed the global gene expression pattern to that of cells not overexpressing *HOTAIR* (Fig. 4c, right). Quantitative PCR with reverse transcription (qRT–PCR) confirmed that *HOTAIR*-induced PRC2 target genes, such as *JAM2*, *PCDH10* and *PCDHB5*, were transcriptionally repressed after *HOTAIR* expression and de-repressed after concomitant PRC2 depletion (Fig. 4d). *HOTAIR*-induced genes were also reversed after PRC2 depletion (Fig. 4d). Of note, many of the genes induced by *HOTAIR* are known positive regulators of cancer metastasis, including *ABL2* (ref. 21), *SNAIL*[22], and laminins[23]. Conversely, overexpression of EZH2 in H16N2 breast cells is known to promote matrix invasion[12], but concomitant depletion of endogenous *HOTAIR* in large measure inhibited the ability of EZH2 to induce matrix invasion (Fig. 4e and Supplementary Fig. 6). Together, these results demonstrate a functional interdependency between *HOTAIR* and PRC2 in promoting cancer invasiveness.

In summary, the cancer transcriptome is more complex than previously believed. In addition to protein-coding genes and microRNAs,

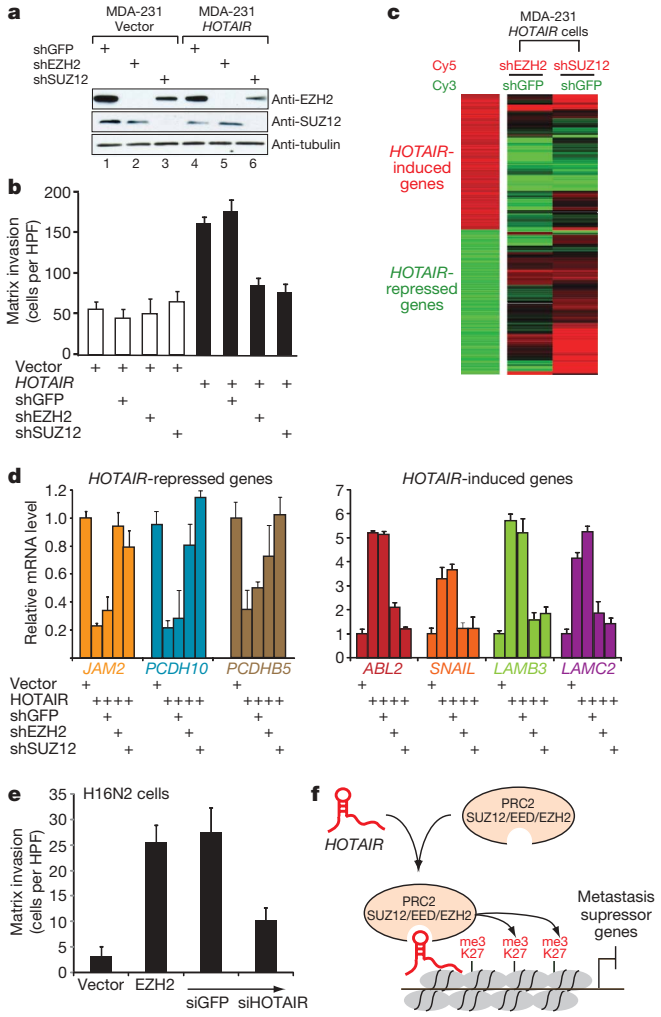

Figure 4 | *HOTAIR*-induced matrix invasion and global gene expression changes requires PRC2. a, Immunoblot of SUZ12 and EZH2 protein levels after transduction of MDA-MB-231 vector or *HOTAIR* cells with retrovirus expressing an shRNA targeting either GFP, EZH2 or SUZ12. **b**, Matrix invasion in vector or *HOTAIR* cells expressing the indicated shRNA. Mean ± s.d. are shown ($n = 3$). **c**, Left, heat map of genes with significant induction (red) or repression (green) after *HOTAIR* expression in the MDA-MB-231 cells. Right, the relative expression of the same gene list in MDA-MB-231 *HOTAIR* cells expressing shEZH2 or shSUZ12 (expressed as a ratio to *HOTAIR* cells expressing shGFP). **d**, qRT–PCR of a representative panel of genes in MDA-MB-231 vector or *HOTAIR* cells also expressing the indicated shRNA (error bars are s.d., $n = 3$). **e**, Matrix invasion in the immortalized H16N2 breast epithelial line expressing vector or EZH2 as well as EZH2-expressing cells transfected with siRNAs targeting GFP or *HOTAIR* (error bars = s.d., $n = 3$). **f**, Working model of the role of *HOTAIR* in breast cancer progression. Selection for increased *HOTAIR* expression in a subset of breast primary tumours leads to a genome-wide retargeting of the PRC2 and H3K27me3 patterns, resulting in gene expression changes that promote tumour metastasis.

dysregulated expression of lincRNAs is probably pervasive in human cancers and can drive cancer development and progression. Notably, the lincRNA *HOTAIR* regulates metastatic progression. *HOTAIR* recruits the PRC2 complex to specific target genes genome-wide, leading to H3K27 trimethylation and epigenetic silencing of metastasis suppressor genes (Fig. 4f). The concept of epigenomic reprogramming by lincRNAs may also be applicable to many other human disease states characterized by aberrant lincRNA expression and chromatin states. *HOTAIR* is normally involved in specifying the chromatin state associated with fibroblasts from anatomically posterior and distal sites. Within the context of cancer cells, ectopic expression of *HOTAIR* seems to re-impose that chromatin state, thereby enabling gene expression programs that are conducive to cell motility and matrix invasion.

The interdependence between *HOTAIR* and PRC2 has therapeutic implications. High levels of *HOTAIR* may identify tumours that are sensitive to small molecule inhibitors of PRC2 (ref. 24). Conversely, tumours that overexpress Polycomb proteins may be sensitive to therapeutic strategies that target endogenous *HOTAIR* or inhibit *HOTAIR*–PRC2 interactions. Understanding the precise molecular mechanisms by which *HOTAIR* regulates PRC2 will be a critical first step in exploring these potential new strategies in cancer therapy.

METHODS SUMMARY

Human material was obtained from Johns Hopkins Hospital and the Netherlands Cancer Institute. Expression of *HOX* transcripts was determined using ultra-high-density HOX tiling arrays[7] and qRT–PCR. Kaplan–Meier analyses of breast cancer patients were as described[13]. We used retroviral transduction to overexpress *HOTAIR* and luciferase, and used siRNA or shRNA to deplete the indicated transcripts. Matrix invasion was measured by the transwell Matrigel assay. We implanted cells in the mammary fat pad of severe combined immunodeficient (SCID) mice, and monitored primary tumour growth and lung metastasis by bioluminescence. Cells were injected into the tail vein of nude mice, and lungs were analysed at 9 weeks to quantify lung colonization *in vivo*. ChIP-chip was performed as described[7] using human whole genome promoter tiling arrays (Roche Nimblegen). Module map and Gene Ontology enrichment analyses were done using Genomica[20].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 August 2009; accepted 26 February 2010.

1. Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **319,** 1787–1789 (2008).
2. The FANTOM Consortium. The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–1563 (2005).
3. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
4. Calin, G. A. *et al.* Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12,** 215–229 (2007).
5. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* **451,** 202–206 (2008).
6. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136,** 629–641 (2009).
7. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
8. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106,** 11667–11672 (2009).
9. Raman, V. *et al.* Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405,** 974–978 (2000).
10. Wu, X. *et al.* HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial-mesenchymal transition. *Cancer Res.* **66,** 9527–9534 (2006).
11. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Natl. Rev.* **6,** 846–856 (2006).
12. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100,** 11606–11611 (2003).
13. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
14. Ma, L., Teruya-Feldstein, J. & Weinberg, R. A. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449,** 682–688 (2007).
15. Novak, P. *et al.* Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res.* **68,** 8616–8625 (2008).
16. Naik, M. U., Naik, T. U., Suckow, A. T., Duncan, M. K. & Naik, U. P. Attenuation of junctional adhesion molecule-A is a contributing factor for breast cancer cell invasion. *Cancer Res.* **68,** 2194–2203 (2008).
17. Fox, B. P. & Kandpal, R. P. Invasiveness of breast carcinoma cells and transcript profile: Eph receptors and ephrin ligands as molecular markers of potential diagnostic and prognostic application. *Biochem. Biophys. Res. Commun.* **318,** 882–892 (2004).
18. Herath, N. I., Doecke, J., Spanevello, M. D., Leggett, B. A. & Boyd, A. W. Epigenetic silencing of EphA1 expression in colorectal cancer is correlated with poor survival. *Br. J. Cancer* **100,** 1095–1102 (2009).
19. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
20. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nature Genet.* **36,** 1090–1098 (2004).

21. Srinivasan, D. & Plattner, R. Activation of Abl tyrosine kinases promotes invasion of aggressive breast cancer cells. *Cancer Res.* **66,** 5648–5655 (2006).
22. Olmeda, D. *et al.* SNAI1 is required for tumor growth and lymph node metastasis of human breast carcinoma MDA-MB-231 cells. *Cancer Res.* **67,** 11721–11731 (2007).
23. Marinkovich, M. P. Tumour microenvironment: laminin 332 in squamous-cell carcinoma. *Natl. Rev.* **7,** 370–380 (2007).
24. Tan, J. *et al.* Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev.* **21,** 1050–1063 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Chen-Tsai, M. Guttman, G. Sen, T. Ridky, P. Khavari, V. Band and Y. Kang for advice and reagents. Supported by National Institutes of Health (NIH), Emerald Foundation, and American Cancer Society (H.Y.C.), Dermatology Foundation (R.A.G., K.C.W. and D.J.W.), Susan Komen Foundation (M.-C.T.), NSF (T.H.), and Department of Defense BCRP (S.S.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

Author Contributions R.A.G. measured lincRNAs in cancer samples and performed all gene transfer and knockdown experiments. R.A.G. and N.S. performed cell growth, invasion, and *in vivo* xenograft assays. R.A.G., K.C.W., M.-C.T. and T.H. performed ChIP-chip studies and analyses. R.A.G., J.L.R. and D.J.W. performed bioinformatic analyses. J.K. performed *in vivo* bioluminescence studies. H.M.H., P.A. and M.J.v.d.V. procured and analysed human tumour samples. Y.W., P.B. and B.K. designed lincRNA Taqman probes and analysed tumour RNAs by qRT–PCR. R.L. and R.B.W. performed *in situ* hybridization studies. R.A.G., N.S., S.S. and H.Y.C. designed the experiments and interpreted the results. R.A.G. and H.Y.C. wrote the paper.

Author Information Microarray data are deposited in Gene Expression Omnibus (GEO) under accession numbers GSE20435 and GSE20737. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.Y.C. (howchang@stanford.edu).

METHODS

Reagents. The MDA-MB-231, SK-BR-3, MCF-10A, MCF-7, HCC1954, T47D and MDA-MB-453 cell lines were obtained from the American Type Culture Collection (ATCC). The H16N2 cell line was a gift from V. Band. pLZRS, pLZRS-luciferase and pSuper Retro-shGFP, -shSUZ12 and -shEZH2 (ref. 25) were obtained from P. Khavari. pLZRS-HOTAIR and pLZRS-EZH2-Flag were constructed by subcloning the full-length human *HOTAIR*[7] or Flag–EZH2–ER fusion protein (representing amino acids 1–751 of EZH2 fused with the murine oestrogen receptor (amino acids 281–599)) into pLZRS using the Gateway cloning system (Invitrogen).

Human materials. Normal breast organoid RNA was prepared as reported[26]. In brief, tissues from reduction mammoplasties performed at Johns Hopkins Hospital were mechanically macerated then digested overnight with hyaluronic acid and collagenase. The terminal ductal units are placed into suspension by this method; they were then isolated by serial filtration. Samples were treated with TRIzol and RNA extracted.

Fresh frozen primary breast tumour specimens were obtained from the Department of Pathology breast tumour bank; specimens were all from patients 45–55 years of age, with oestrogen receptor expression by immunohistochemistry as performed during routine tumour staging at diagnosis, for uniformity of samples.

Metastatic breast carcinoma samples were obtained from the Rapid Autopsy Program at Johns Hopkins Hospital[27]. All specimens were snap-frozen at time of autopsy and stored at −80 °C. Twenty 20-μm sections were obtained from metastasis to the liver (for uniformity of samples) and embedded in OCT. These slices were macerated by use of the BioMasher centrifugal sample preparation device (Cartagen), with 350 μl of lysis buffer from the Qiagen RNeasy Mini Extraction kit. RNA extraction was completed with the flow-through from the BioMasher, as per the commercial protocol.

***HOTAIR* expression and survival/metastasis analysis of primary breast tumours.** The database of 295 breast cancer patients from the Netherlands Cancer Institute with detailed clinical and gene expression data was used[13]. Clinical data are available at http://microarray-pubs.stanford.edu/wound_NKI, http://www.rii.com/publications, or http://microarrays.nki.nl. RNA from 132 primary breast tumours from the NKI 295 cohort was isolated along with RNA from normal breast organoid cultures ($n = 6$). *HOTAIR* and *GAPDH* were measured by qRT–PCR. *HOTAIR* values were normalized to *GAPDH* and expressed relative to pooled normal *HOTAIR* RNA levels. For both univariate and multivariate analysis, the expression of *HOTAIR* was treated as a binary variable divided into 'high' and 'low' *HOTAIR* expression. To determine the criteria for high *HOTAIR* expression, the minimum relative level of *HOTAIR* seen in six metastatic breast cancer samples (see Fig. 1c and accompanying methods) was determined (≥125 above normal). By this criteria, 44 primary breast tumours were categorized as high, and 88 were labelled as low, out of 132 tumours. For statistical analysis, overall survival was defined by death from any cause. Distant metastasis-free probability was defined by a distant metastasis as the first recurrence event. Kaplan–Meier survival curves were compared by the Cox–Mantel log-rank test in Winstat (R. Fitch software). Multivariate analysis by the Cox proportional hazard method was done using SPSS 15.0 (SPSS)

RNA expression analysis. qRT–PCR: total RNA from cells was extracted using TRIzol and the RNeasy mini kit (Qiagen). RNA levels (starting with 50–100 ng per reaction) for a specific gene (primer set sequences listed in Supplementary Table 4) were measured using the Brilliant SYBR Green II qRT–PCR kit (Strategene) according to manufacturer instructions. All samples were normalized to *GAPDH*.

HOX tiling array: RNA samples (primary or metastatic breast carcinoma in channel in Cy5 channel and normal breast organoid RNA representing a pool of six unique samples in Cy3 channel) were labelled and hybridized to a custom human *HOX* tiling array with 50-base-pair resolution (Roche Nimblegen) as described[7]. For each sample, robust multichip average (RMA) normalized intensity values for previously defined peaks encoding *HOX*-coding-gene exons (as defined in version HG17) and *HOX* lincRNAs (as defined previously[7]) were determined relative to normal. Unsupervised hierarchical clustering was performed by CLUSTER[28].

Microarray: total RNA from cells was extracted using TRIzol and the RNeasy mini kit (Qiagen) and hybridized to Stanford human oligonucleotide (HEEBO) arrays as described[29]. Data analysis was done using CLUSTER[28].

Gene transfer experiments. Retrovirus was generated using amphotrophic phoenix cells and used to infect target cells as described[30]. For LZRS vector, *HOTAIR*, EZH2–ER, and firefly luciferase, no further selection was done after infection. For pRetro-Super-shGFP, -shSUZ12 and -shEZH2, target cells were selected using puromycin (0.5 μg ml^{-1}). Many of the epigenetic changes due to *HOTAIR* expression were only seen after several cell passages; thus all experiments post-*HOTAIR* transduction were done after passage 10.

Non-radioactive *in situ* hybridization of paraffin sections. Digoxigenin (DIG)-labelled sense and antisense RNA probes were generated by PCR amplification of T7 promoter incorporated into the primers. *In vitro* transcription was performed with DIG RNA labelling kit and T7 polymerase according to the manufacturer's protocol (Roche Diagnostics). Sections (5-μm thick) were cut from the paraffin blocks, deparaffinized in xylene, and hydrated in graded concentrations of ethanol for 5 min each. Sections were incubated with 1% hydrogen peroxide, followed by digestion in 10 μg ml^{-1} proteinase K at 37 °C for 30 min. Sections were hybridized overnight at 55 °C with either sense or antisense riboprobes at 200 ng ml^{-1} dilution in mRNA hybridization buffer (Chemicon). The next day, sections were washed in 2×SSC and incubated with 1:35 dilution of RNase A cocktail (Ambion) in 2×SSC for 30 min at 37 °C. Next, sections were stringently washed twice in 2×SSC/50% formamide, followed by one wash in 0.08×SSC at 55 °C. Biotin-blocking reagents (Dako) were applied to the section to block the endogenous biotin. For signal amplification, a horseradish peroxidase (HRP)-conjugated sheep anti-DIG antibody (Roche) was used to catalyse the deposition of biotinyl-tyramide, followed by secondary streptavidin complex (GenPoint kit; Dako). The final signal was developed with DAB (GenPoint kit; Dako), and the tissues were counter-stained in haematoxylin for 30 s.

RNA interference. RNA interference for *HOTAIR* was done as described[7]. In brief, cells were transfected with 50 nM siRNAs targeting *HOTAIR* (siHOTAIR-1, 5′-GAACGGGAGUACAGAGAGAUU-3′; siHOTAIR-2, 5′-CCACAUGAACGC CCAGAGAUU-3′; siHOTAIR-3, 5′-UAACAAGACCAGAGAGCUGUU-3′) or siGFP (5′-CUACAACAGCCACAACGUCdTdT-3′) using Lipofectamine 2000 (Invitrogen) as per the manufacturer's direction. Total RNA was collected 72 h later for qRT–PCR analysis.

RNA interference of EZH2 and SUZ12 was done by infecting target cells with retrovirus expressing shEZH2, shSUZ12 and shGFP as described[25]. To confirm knockdown, protein lysates were resolved on 10% SDS–PAGE followed by immunoblot analysis as described[30] using anti-SUZ12 (Abcam), anti-EZH2 (Upstate) and anti-tubulin (Santa Cruz).

Matrigel invasion assay and cell proliferation assay. The matrigel invasion assay was done using the Biocoat Matrigel Invasion Chamber from Becton Dickson according to manufacturer protocol. In brief, 5×10^4 cells were plated in the upper chamber in serum-free media. The bottom chamber contained DMEM media with 10% FBS. After 24–48 h, the bottom of the chamber insert was fixed and stained with Diff-Quick stain. Cells on the stained membrane were counted under a dissecting microscope. Each membrane was divided into four quadrants and an average from all four quadrants was calculated. Each matrigel invasion assay was at least done in biological triplicates. For invasion assays in the H16N2 cell line using EZH2–ER, all experiments (both vector and with EZH2–ER) were done in the presence of 500 nM oestradiol.

For cell proliferation assays, 1×10^3 cells were plated in quadruplicate in 96-well plates and cell number was calculated using the MTT assay (Roche).

Soft agar colony formation assay. Soft agar assays were constructed in 6-well plates. The base layer of each well consisted of 2 ml with final concentrations of 1× media (RPMI (HCC1954), McCoy's Media (SKBR3), or DMEM (MDA-MB-231) plus 10% or 2% heat-inactivated FBS (Invitrogen)) and 0.6% low melting point agarose. Plates were chilled at 4 °C until solid. Upon this, a 1-ml growth agar layer was poured, consisting of 1×10^4 cells (infected with either LZRS-HOTAIR or LZRS vector as described earlier) suspended in 1× media and 0.3% low melting point agarose. Plates were again chilled at 4 °C until the growth layer congealed. A further 1 ml of 1× media without agarose was added on top of the growth layer on day 0 and again on day 14 of growth. Cells were allowed to grow at 37 °C for 1 month and total colonies were counted (>200 μm in diameter for MDA-MB-231; >50 μm in diameter for HCC1954 and SKBR3). Assays were repeated a total of three times. Results were statistically analysed by paired *t*-test using the PRISM Graphpad program.

Mammary fat pad xenografts. Six-week-old female SCID beige mice were purchased from Charles River laboratories, housed at the animal care facility at Stanford University Medical Center and kept under standard temperature, humidity and timed lighting conditions and provided mouse chow and water *ad libitum*. MDA-MB-231-Luc or MDA-MB-231-Luc tumour cells transduced with *HOTAIR* were injected directly into the mammary fat pad of the mice semi-orthotopically ($n = 10$ each) in 0.05 ml of sterile DMEM (2,500,000 cells per animal).

Mouse tail-vein assay. Female athymic nude mice were used. Two-million MDA-MB-231 HOTAIR-luciferase or vector-luciferase cells in 0.2 ml PBS were injected by the tail vein into individual mice (18 for each cell line). Mice were observed generally for signs of illness weekly for the length of the experiment. The lungs were excised and weighed fresh, then bisected. Half was fixed in formalin overnight then embedded in paraffin, from which sections were made and stained with haematoxylin and eosin by our pathology consultation service. These slides were examined for the presence of micrometastases, which were

counted in three low-power (×5) fields per specimen. The other half of the tumour was fast-frozen into OCT and stored at −80 °C. RNA was extracted by the TRIzol protocol from ten sections, 20-μm thick each, obtained from the frozen sections. RT–PCR confirmed expression of *HOTAIR* RNA in lungs bearing micrometastases of MDA-MB-231 HOTAIR cells at the end of the experiment.

Bioluminescence imaging. Mice received luciferin (300 mg kg^{-1}, 10 min before imaging) and were anaesthetized (3% isoflurane) and imaged in an IVIS spectrum imaging system (Xenogen, part of Caliper Life Sciences). Images were analysed with Living Image software (Xenogen, part of Caliper Life Sciences). Bioluminescent flux (photons s^{-1} sr^{-1} cm^{-2}) was determined for the primary tumours or lungs (upper abdomen region of interest).

ChIP-chip. ChIP-chip experiments were done as previously described[7]. Each experiment was done in biological triplicate. The following antibodies were used: anti-H3K27me3 (Abcam), anti-SUZ12 (Abcam) and anti-EZH2 (Upstate). Immunoprecipitated DNA was amplified using the Whole Genome Amplification kit (Sigma) based on the manufacturer's protocol. Amplified and labelled DNA was hybridized to the HG18 whole genome two array promoter set from Roche Nimblegen. Probe labelling, hybridization, and data extraction and analysis were performed using Roche Nimblegen protocols. The relative ratio of *HOTAIR* to vector was calculated for each promoter peak by extracting the normalized (over input) intensity values for promoter peaks showing peaks with an FDR score ≤ 0.2 in either vector or *HOTAIR* cells. These values were weighted to determine the significance of the relative ratio: using Cluster[28], only those promoters with a consistent relative ratio (*HOTAIR*/vector) ≥1.5-fold or ≤0.5-fold in two out of the three ChIP were selected and displayed in TreeView. Selected ChIP-chip results were confirmed by PCR using the Lightcycler 480 SYBR Green I kit (see Supplementary Table 5 for primer sequences).

TaqMan real-time PCR assays. A panel of 96 TaqMan real-time PCR HOX assays (Supplementary Table 6) was developed targeting 43 *HOX* lincRNAs and 39 HOX transcription factors across the four *HOX* loci. Two housekeeping genes (*ACTB* and *PPIA*) were also included in this panel in triplicates as endogenous controls for normalization between samples. The transcript specificity and genome specificity of all TaqMan assays were verified using a position-specific alignment matrix to predict potential cross-reactivity between designed assays and genome-wide non-target transcripts or genomic sequences. Using this HOX assay panel we profiled 88 total RNA samples from a cohort of five normal breast organoids, 78 primary breast tumours (from the NKI 295 cohort) and five metastatic breast tumours. cDNAs were generated from 30 ng total RNA using the High Capacity cDNA Reverse Transcription Kit (Life Technologies). The resulting cDNA was subjected to a 14-cycle PCR amplification followed by real-time PCR reaction using the manufacturer's TaqMan PreAmp Master Mix Kit Protocol (Life Technologies). Four replicates were run for each gene for each sample in a 384-well format plate on a 7900HT Fast Real-Time PCR System (Life Technologies). Between the two measured endogenous control genes (*PPIA* and *ACTB*), we chose *PPIA* for normalization across different samples based on the fact that this gene showed the most relatively constant expression in different breast carcinomas (data not shown).

Gene set analysis. For gene set enrichment analysis, gene sets from fifteen different H3K27, SUZ12 or EZH2 global occupancy lists from the indicated cell lineages were procured (see Supplementary Table 3 for references and platforms). Pattern matching between the 854-gene set with increased PRC2 occupancy (Supplementary Table 7) and these 15 gene sets were visualized using CLUSTER and TreeView. The significance of enrichment between these gene sets was calculated using module map analysis implemented in Genomica[20] (corrected for multiple hypotheses using FDR).

25. Sen, G. L., Webster, D. E., Barragan, D. I., Chang, H. Y. & Khavari, P. A. Control of differentiation in a self-renewing mammalian tissue by the histone demethylase JMJD3. *Genes Dev.* **22,** 1865–1870 (2008).
26. Bergstraesser, L. M. & Weitzman, S. A. Culture of normal and malignant primary human mammary epithelial cells in a physiological manner simulates *in vivo* growth patterns and allows discrimination of cell type. *Cancer Res.* **53,** 2644–2654 (1993).
27. Wu, J. M. *et al.* Heterogeneity of breast cancer metastases: comparison of therapeutic target expression and promoter methylation between primary tumors and their multifocal metastases. *Clin. Cancer Res.* **14,** 1938–1946 (2008).
28. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
29. Rinn, J. L., Bondre, C., Gladstone, H. B., Brown, P. O. & Chang, H. Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* **2,** e119 (2006).
30. Adler, A. S. *et al.* Genetic regulators of large-scale transcriptional signatures in cancer. *Nature Genet.* **38,** 421–430 (2006).