Genome Biology

RESEARCH

Open Access

Endometrial tumorigenesis involves epigenetic plasticity demarcating non-coding somatic mutations and 3D-genome alterations

Sebastian Gregoricchio[1*], Aleksandar Kojic[1], Marlous Hoogstraat[1,2], Karianne Schuurman[1], Suzan Stelloo[3], Tesa M. Severson[1,2], Tracy A. O'Mara[4], Marjolein Droog[1], Abhishek A. Singh[1], Dylan M. Glubb[4], Lodewyk F. A. Wessels[2], Michiel Vermeulen[3,5], Flora E. van Leeuwen[6] and Wilbert Zwart[1,7*]

*Correspondence: s.gregoricchio@nki.nl; w.zwart@nki.nl

[1] Division of Oncogenomics, Oncode Institute, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands
Full list of author information is available at the end of the article

Abstract

Background: The incidence and mortality of endometrial cancer (EC) is on the rise. Eighty-five percent of ECs depend on estrogen receptor alpha (ERα) for proliferation, but little is known about its transcriptional regulation in these tumors.

Results: We generate epigenomics, transcriptomics, and Hi-C datastreams in healthy and tumor endometrial tissues, identifying robust ERα reprogramming and profound alterations in 3D genome organization that lead to a gain of tumor-specific enhancer activity during EC development. Integration with endometrial cancer risk single-nucleotide polymorphisms and whole-genome sequencing data from primary tumors and metastatic samples reveals a striking enrichment of risk variants and non-coding somatic mutations at tumor-enriched ERα sites. Through machine learning-based predictions and interaction proteomics analyses, we identify an enhancer mutation which alters 3D genome conformation, impairing recruitment of the transcriptional repressor EHMT2/G9a/KMT1C, thereby alleviating transcriptional repression of *ESR1* in EC.

Conclusions: In summary, we identify a complex genomic-epigenomic interplay in EC development and progression, altering 3D genome organization to enhance expression of the critical driver ERα.

Keywords: Endometrial cancer, Estrogen receptor, Epigenetic plasticity in tumor development, Gene regulation, 3D genome organization, Non-coding somatic mutations

Background

Endometrial cancer (EC) is the second most-common gynecological cancer with over 417,000 new cases and 97,000 deaths in 2020, with global incidence rates increasing every year [1]. Surprisingly, and in stark contrast to most other cancers, survival rates for endometrial cancer are decreasing [1, 2]. This gradual deterioration of EC survival

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

is likely related to the relatively understudied nature of the disease, with many molecular mechanisms driving tumor development and progression remaining largely elusive.

Endometrial tissue is under tight endocrine control, and this remains the case upon tumorigenesis. The majority of endometrial cancers (~85%) are classified as low grade endometrioid tumors, expressing estrogen receptor alpha (ERα) and progesterone receptor (PR) [3–5]. PR agonists are used in the treatment of endometrial tumors as alternative to chemotherapy and hysterectomy to retain uterine function in young patients [6]. Furthermore, the estrogen competitive ERα antagonist tamoxifen is also prescribed to patients with endometrial cancer, often alternated with PR agonists (reviewed in ref [7]). Thus, while PR acts in a tumor suppressive manner, ERα serves as driver of tumor progression which is therapeutically blocked in endometrial cancer care.

ERα is an enhancer-acting transcription factor, regulating expression of its responsive genes through long-range 3D genome interactions [8, 9]. To date, most literature on promoter-enhancer communication and ERα activity has been focused on breast cancer [9] and is far less understood in endometrial tumors. Between breast cancer and endometrial cancer, substantial differences are observed in ERα chromatin binding profiles [10–12], highlighting the highly context-dependent nature of ERα genomic action. Paradoxically, while tamoxifen can be prescribed in the treatment of endometrial cancer, its use in the treatment of breast cancer is also reported as risk factor for endometrial tumorigenesis [13]. Previously, we showed that tamoxifen treatment in endometrial tissue reprograms the ERα chromatin interaction landscape, phenocopying profiles found in breast cancer cells, driving endometrial tumor growth [14]. These observations highlighted that the endometrial cancer epigenome is not fixed, but rather dynamically affected by oncogenic factors.

Both ERα in breast tissue [15] and androgen receptor (AR) in prostate tissue [16–18] undergo extensive reprogramming throughout the genome during tumorigenesis. In case of prostate cancer, both somatic mutations and risk single-nucleotide polymorphisms (SNPs) are enriched at tumor-gained AR sites, with only a small fraction of which causally affecting transcriptional output [17]. In endometrial cancer, such studies have thus far not been reported, and functional interplay between somatic mutations, epigenetic alterations, and the impact on 3D genome organization remain fully elusive.

Here, we investigated the plasticity of ERα genomic action upon endometrial tumor development, by comparing the epigenome of primary human healthy and tumor endometrial tissues (ERα and H3K27ac ChIP-seq) and the effects on 3D genome organization (Hi-C, 4C-seq, and H3K27ac HiChIP). Secondly, we studied the crosstalk between epigenetic alterations and somatic variant events (WGS) that are associated with EC progression. We identified substantial epigenetic reprogramming upon tumorigenesis that results in the ERα re-localization at tumor-specific regions throughout the genome, which coincided with the occurrence of somatic variants in metastatic samples. In particular, we discovered an EC-specific *ESR1* enhancer which is selectively found mutated in metastatic EC. In vitro analyses show diminished capacity to bind lysine methyltransferase G9a/EHMT2/KMT1C to this region when mutated, and perturbation of G9a expression enhanced ERα expression in cell line studies. Cumulatively, we show that non-coding mutations in endometrial cancer may have direct pro-tumorigenic potential,

through a tight interplay with epigenetic alterations and changes in the 3D genome structure.

Results

ERα enhancer plasticity in endometrial tumorigenesis

To investigate the role of ERα signaling in endometrial tumor development, we first sought to assess how the ERα cistrome differs between endometrial tumors and healthy endometrial tissue. We collected five fresh-frozen endometrial tumors from post-menopausal patients that did not receive any estrogen receptor targeting for cancer therapy, as well as four fresh-frozen samples from post-menopausal women with pathologically normal endometrial tissue (Fig. 1a and Additional file 1: Fig. S1a, for clinicopathological characteristics see Additional file 2: Table S1). All samples were subjected to ERα and H3K27 acetylation (H3K27ac) ChIP-seq in order to identify active regulatory elements that are under ERα control. On average, we identified 20,815 ERα peaks (range 1399–51,888) and 44,824 H3K27ac peaks (range 8791–71,898) in our samples. The ChIP-seq library size was comparable among the different samples for both targets (Additional file 1: Fig. S1b). Interestingly, the number of ERα peaks was higher in tumor samples compared to healthy tissue (Mann-Whitney $P = 0.063$, Fig. 1b) while the global number of H3K27ac peaks for these samples was comparable between both groups (Mann-Whitney $P = 0.29$, Fig. 1c). To assess data quality and reproducibility, we performed principal component analysis (PCA) on the peak intensities which revealed a clustering of the samples by ChIP target (first component) and tissue type (second component) (Additional file 1: Fig. S1c). Furthermore, given the relatively small sample size, we tested for over-training and performed unsupervised permutation clustering analyses that showed a clear separation of the two tissues types, except for the sample T33 showing mixed features, between normal and tumor tissues (Additional file 1: Fig. S1d), possibly due to a lower tumor cell representation in those sections that were used for ChIP-seq studies. Similarly, sample T119—even though clustering with the other tumor

(See figure on next page.)

Fig. 1 ERα cistrome changes upon tumorigenesis. **a** Schematic workflow of the multi-omics approach applied in this work. Boxplot depicting the distribution of peak number detected by ERα (**b**) or H3K27ac (**c**) ChIP-seq in healthy (blue) or tumor (orange) endometrial tissues. **d** Venn diagram of the overlap between consensus ERα binding sites detected in healthy (blue) and tumor (orange) tissues. Consensus peaks of each group have been defined as peaks present in at least 75% of the samples belonging to that specific group. **e** MA plot of differential binding analyses performed on the consensus ERα ChIP-seq peaks (healthy vs tumor). Differential binding sites (FDR ≤ 0.05 and $|\log_2(\text{FoldChange})| \geq 1$) are highlighted in pink. Tornado plot of the ERα (**f**) or H3K27ac (**g**) ChIP-seq signal at tumor-depleted (blue, upper blocks) or tumor-enriched (orange, lower blocks) ERα consensus peaks for all the 4 healthy (blue, left heatmaps) and 5 tumor (orange, right heatmaps) endometrial tissues. On the top of each heatmap is plotted the average density signal for each region group. **h** Stacked bar plot depicting the genomic localization frequency of tumor-depleted and tumor-enriched ERα ChIP-seq consensus peaks. **i** Stacked bar plot displaying the distance to associated gene TSS frequency distribution of tumor-depleted and tumor-enriched ERα ChIP-seq consensus peaks. **j** Wordcloud of the top-50 enriched motifs at tumor-depleted (blues) and tumor-enriched (oranges) ERα consensus peaks. Size and color intensity are proportional to the $-\log_{10}(E\text{-value})$. **k** The upper part of the heatmap shows the individual tumor-depleted (blues) and tumor-enriched (oranges) ERα consensus peaks colored by signal intensity. On the lower part, each black bar represents an overlapping peak of ChIP-seq data of several targets publicly available for the Ishikawa endometrial cancer cell line. **l** Heatmap of the percentage of tumor-depleted or tumor-enriched ERα consensus peaks overlapping with each ChIP-seq target in Ishikawa cells from (**k**). Ranking is performed by descending number of overlaps in tumor-depleted peaks

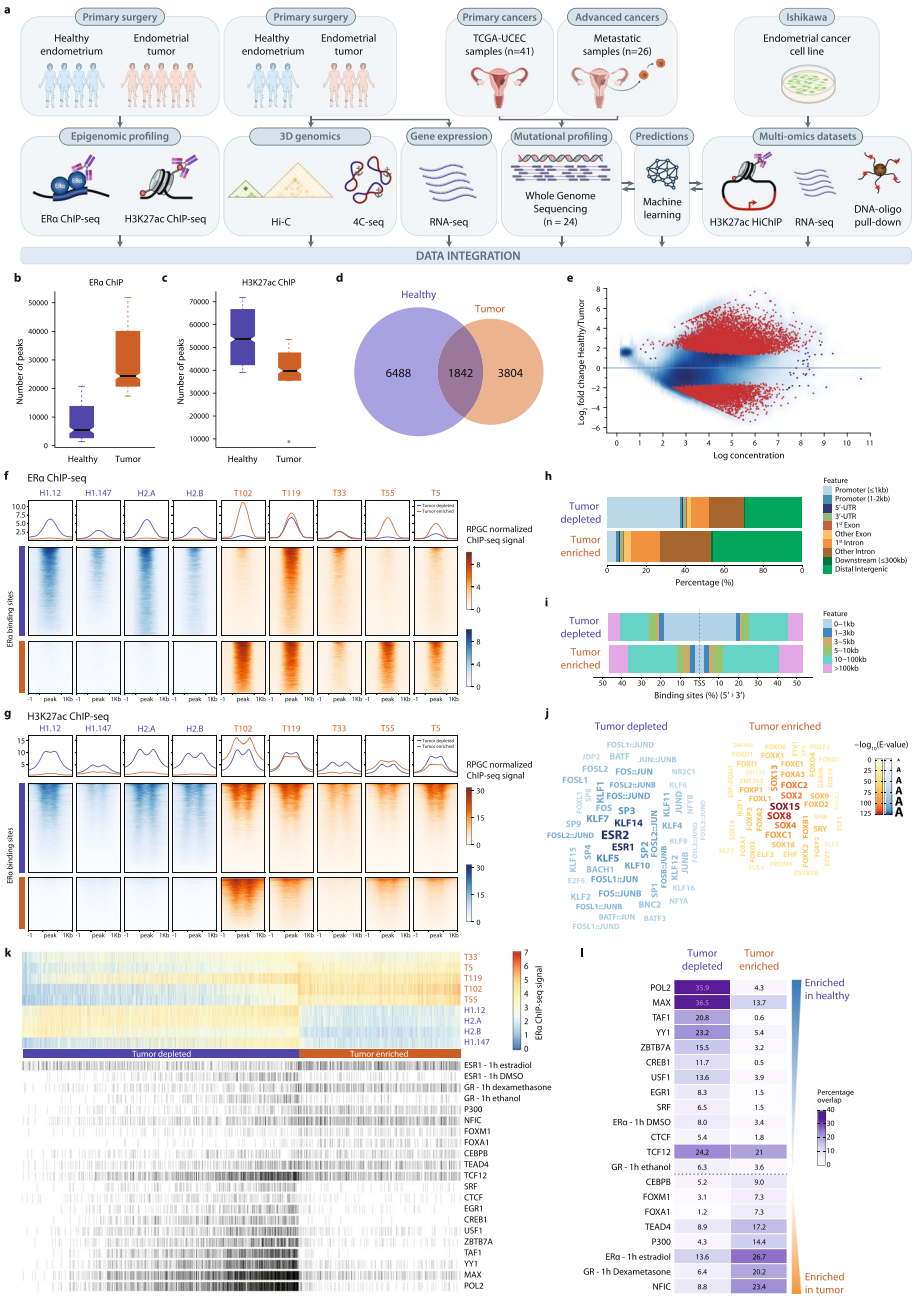

Fig. 1 (See legend on previous page.)

samples (Additional file 1: Fig. S1d)—displays ERα signal at the tumor-depleted regions (Fig. 1f). While tissues were selected based on a high tumor percentage (Additional file 2: Table S1), we cannot exclude that further sections used in our analyses may have a relatively larger non-tumor cell fraction. On the other hand, as we cannot exclude that this phenomenon may represent inter-patient epigenetic heterogeneity, we therefore decided to retain all samples for further analyses.

Differential binding analysis (*DiffBind* [19]) for ERα between tumor and normal samples revealed 10,292 differentially bound genomic locations (FDR $\leq$ 0.05 and

$|\log_2(\text{FoldChange})| \geq 1$) (Fig. 1d–e, Additional file 1: Fig. S1f, and Additional file 1: Fig. S2), of which 6488 ERα sites were lost and 3804 ERα sites gained in tumor samples, hereafter referred to as "tumor-depleted" (blue) and "tumor-enriched" (orange) ERα binding sites, respectively. Tornado plots depicting normalized ERα ChIP-seq signal for all 10,292 differential-enriched ERα sites (Fig. 1f), and label-swap permutation differential binding analyses (Additional file 1: Fig. S1e), confirmed the selective enrichment of signals between samples, with high reproducibility. As ERα is a transcription factor that acts predominantly by occupying enhancers [8, 9], we next analyzed the differential ERα sites for presence of H3K27ac, inferring regulatory element activity. Interestingly, persistent activity in both tissue states was observed at tumor-depleted ERα sites, while tumor-enriched ERα sites showed a marked increase of activity upon tumorigenesis (Fig. 1g and Additional file 1: Fig. S2).

Collectively, these data suggest a selective gain of regulatory element potential upon tumorigenesis, marked by dynamically altered ERα sites.

Genomic distributions and transcription complex composition at differential ERα sites

To further investigate the nature of the ERα regions differentially bound upon tumorigenesis, we first inspected their genome-wide distribution with respect to different genomic elements. In agreement with previous reports from us (tumors [10, 14]) and others (cell lines [11]), tumor-enriched ERα sites were mostly found at distal intergenic and intronic regions (Fig. 1h) and at larger distance from the associated transcription start site (TSS) (Fig. 1i). However, tumor-depleted ERα sites, selectively occupied by ERα in healthy endometrial tissue, display a strong enrichment for promoter regions of genes involved in cell cycle regulation and DNA damage processes (Additional file 1: Fig. S1g). To determine whether loss of ERα binding at these promoters leads to transcriptional deregulation of the associated genes in tumor samples, we performed RNA sequencing (RNA-seq) in 3 healthy and 3 tumor endometrial tissues from post-menopausal patients that did not receive any prior estrogen receptor targeting for cancer therapy (Additional file 2: Table S1). Principle component analyses display a clear separation of normal and tumor tissues with a higher variability for tumor samples than healthy ones (Additional file 1: Fig. S3a). Differential expression analyses identified 1945 genes to be downregulated and 2392 to be upregulated in tumors, compared to healthy tissue (Additional file 1: Fig. S3b). Next, we analyzed the expression of the tumor-depleted ERα promoter-bound genes included in the previously identified over-represented biological processes (Additional file 1: Fig. S1g and Additional file 1: Fig. S3c). This investigation confirmed our former results highlighting a significant deregulation of cell division and DNA-repair related genes (Additional file 1: Fig. S3c). Furthermore, results from gene set enrichment analyses (GSEA) for the "hallmarks" (H) dataset also corroborated with these findings in a more global manner (Additional file 1: Fig. S3d). Of note, we observed that differentially expressed genes associated with tumor-enriched ERα binding sites were mostly upregulated in tumor samples compared to genes associated with tumor-depleted ERα binding sites (Fisher's exact $P < 2.2 \times 10^{-16}$, OR 6.8, 95% CI 4.35–10.84; Additional file 1: Fig. S3e–f).

Consensus motifs for ERα/β (ESR1/2) and AP-1 (JUN, FOS) binding sites were specifically enriched at tumor-depleted ERα regions, whereas tumor-enriched regions

demonstrated an overabundance of SOX and FOX transcription factor family member motifs (Fig. 1j). Interestingly, tumor-enriched ERα binding sites, despite the consensus ERα/ESR1 motif being poorly detected, do harbor this motif in around 85% of the cases in a measure comparable to the tumor-depleted sites (Additional file 1: Fig. S1h), albeit highly degenerated as compared to tumor-depleted or common ones (Additional file 1: Fig. S1i). Differential motif enrichment analyses led us to hypothesize that the ERα transcription complex may differ between tumor-enriched and tumor-depleted sites. To test this hypothesis, we next integrated our tumor-enriched and tumor-depleted sites with publicly available ChIP-seq data from the Ishikawa endometrial cancer cell line (Fig. 1k and Additional file 3: Table S2), revealing striking differences between transcription and epigenetic factors on chromatin occupancy for the two ERα binding site categories. Interestingly, ERα chromatin binding in hormone-deprived cells showed a significant overlap with the tumor-depleted ERα peaks (adjusted Fisher's exact $P = 3.15 \times 10^{-21}$), while ERα peaks from estradiol-stimulated cells showed significant overlap with the tumor-enriched ERα peaks (adjusted Fisher's exact $P = 4.60 \times 10^{-59}$, Fig. 1l). In addition, both FOXM1 and FOXA1 binding sites in Ishikawa showed significant overlaps with our tumor enriched regions (adjusted Fisher's exact $P = 3.37 \times 10^{-21}$ and $P = 6.25 \times 10^{-57}$, respectively, Fig. 1l). These findings were successfully confirmed by *GIGGLE* [20] analyses (testing for overlap in a large repository of thousands of publicly available ChIP-seq datasets) at differential ERα binding sites (Additional file 1: Fig. S1j) implicating a strongly divergent transcription complex composition between tumor-enriched and tumor-depleted ERα sites.

Altogether, these data suggest prominent enhancer plasticity in endometrial tumorigenesis, with ERα relocating to alternatively engaged and activated non-canonical enhancers in endometrial cancer.

3D high-order chromatin structures are altered in tumors

Enhancer regions can modulate gene transcription through interactions with promoters in 3D genomic space [21]. With the observed plasticity of ERα chromatin profiles in tumorigenesis, we hypothesized that these epigenetic alterations were accompanied by reorganization of the 3D chromatin structure. To test this hypothesis, we performed high-throughput chromosome conformation capture (Hi-C) analyses on healthy endometrial tissue ($n = 3$) and primary tumors ($n = 3$) derived from post-menopausal women that did not receive any previous systemic therapy (Fig. 2a, for clinicopathological features see Additional file 2: Table S1). Hi-C is particularly suitable to study translocations in cancer [22], which were not observed in any of the tumor samples we analyzed (Additional file 1: Fig. S4a). We did however observe a distinct clustering of the healthy tissues from the tumor specimens, based on 3D genome organization (Additional file 1: Fig. S4b) and independent of Hi-C contact quality bias, intrinsic to the tissue type (Additional file 1: Fig. S4c). In addition, tumor samples displayed a higher Hi-C contact heterogeneity in comparison to the more correlated healthy tissues (Additional file 1: Fig. S4b). Next, we quantified the relative Hi-C contact probability (RCP) as function of the contact distances. Interestingly, tumors displayed an increased probability of shorter-range interactions (<2 Mb) relative to healthy tissue, which is concomitant with a loss of longer-range chromatin contacts (Fig. 2b–d and Additional file 1: Fig. S4 d–e),

highlighting a major 3D genome reorganization in tumor samples independently of the inter-individual variability. Furthermore, analyzing the distribution of the loop distances (Fig. 2e and Additional file 1: Fig. S5), we found that loops whose anchors include tumor-enriched ERα binding sites tend to be shorter in tumor tissues than in healthy tissues. On the other hand, tumor-depleted and shared ERα binding sites do not show differences in loop length between two tissues stages. We therefore investigate the distribution of the Hi-C contacts as function of the distance from different categories of ERα binding sites, by performing paired-end spatial chromatin analysis (PE-SCAn [23]) between healthy and tumor tissues (Fig. 2f). We observed that shared and tumor-depleted ERα binding sites lose chromatin interaction in their surrounding regions in tumor tissues, whereas tumor-enriched binding sites show an increase in relatively shorter-range interactions (Fig. 2f). We next wondered whether the loop shortening, and increased chromatin contacts, at tumor-enriched binding sites in tumor samples was associated with changes in chromatin compartmentalization. Analysis of compartment polarization (Fig. 2b, g–i and Additional file 1: Fig. S4f–g) revealed decreased compartment strength in tumor samples compared to healthy tissues. In particular, we observed that A compartments (euchromatin) were more robust in healthy samples, while B compartments (heterochromatin) were slightly stronger in tumor tissues (Fig. 2h and Additional file 1: Fig. S4g). Overall, around one third of compartments (A: 565/1431 (39.5%); B: 440/1306 (33.6%)) underwent class switching in tumor samples (Fig. 2i). These compartment switches are associated to coherent differences in the expression of genes included in these compartments, when comparing tumors with normal tissue (Additional file 1: Fig. S4h): genes found located in A-to-B are downregulated, while genes located in B-to-A regions are upregulated, in tumors relative to healthy tissue. To investigate whether this compartment class switch was occurring specifically in compartments where ERα was

(See figure on next page.)

Fig. 2 3D genome landscape is remodeled during tumorigenesis. **a** Schematic of the Hi-C library preparation starting from fresh frozen tissues. Ten 30-µm-thick slices of flash-frozen tissue are cross-linked 25 min in 2% formaldehyde. Then, electronically homogenized tissues are filtered using a 75-µm cell strainer and subjected to Hi-C library preparation. Hi-C libraries are then sequenced and resulting reads are analyzed by the *snHiC* [24] pipeline. **b** On the left, 40-kb-resolution matrices of the average Hi-C score in tumor ($n = 3$) and healthy endometrial tissues ($n = 3$) at chromosome 6. On the right, differences of the scores shown on the left side (tumor − healthy). **c** Relative Hi-C contact probability (RCP) as function of the distance for each individual sample (40-kb resolution). **d** Violin plot of the distribution of the short-range (<2 Mb) over long-rage (>2 Mb) Hi-C contacts ratio in each individual sample (40-kb resolution). For each comparison, the Wilcoxon's test *P* value is indicated. **e** Distribution of the Hi-C loop distance (10-kb resolution) for loops overlapping with tumor-depleted, tumor-enriched, or shared ERα consensus peaks in combined healthy (blue) or tumor (orange) endometrial tissues. **f** PE-SCAn results in healthy and tumor tissues at common, tumor-depleted, and tumor-enriched ERα binding sites. For each panel, top and middle rows depict the 3D and 2D, respectively, representation of the Hi-C contact frequency distribution in healthy (left) and tumor (right) tissues; lower row shows the difference of Hi-C contact frequencies between tumor (orange) and healthy (blue) tissue scores. **g** Compartment polarization ratio (100-kb resolution), defined as (AA + BB)/(AB + BA), for each individual sample. **h** *Top*: saddle plot of A/B compartments interactions as computed in **g** for each individual sample. *Bottom*: difference of saddle-score with the reference H.005.A2 (healthy tissue), where orange indicates a higher score in tumor samples while purple indicates a higher score in healthy tissues. **i** Sankey plot depicting the proportion of compartment state transition (100-kb resolution) of healthy compartments (left) upon tumorigenesis (right). **j** Upset plot showing the overlaps between different compartment transition states (100-kb resolution) in healthy compared to tumor as in **g** (A-to-A, B-to-B, A-to-B, B-to-A) and different categories of ERα consensus peaks. **k** Stacked bar plot of the proportion of compartment transition status as in **g** for individual genomic bins overlapping with only one of the different categories of ERα consensus peaks

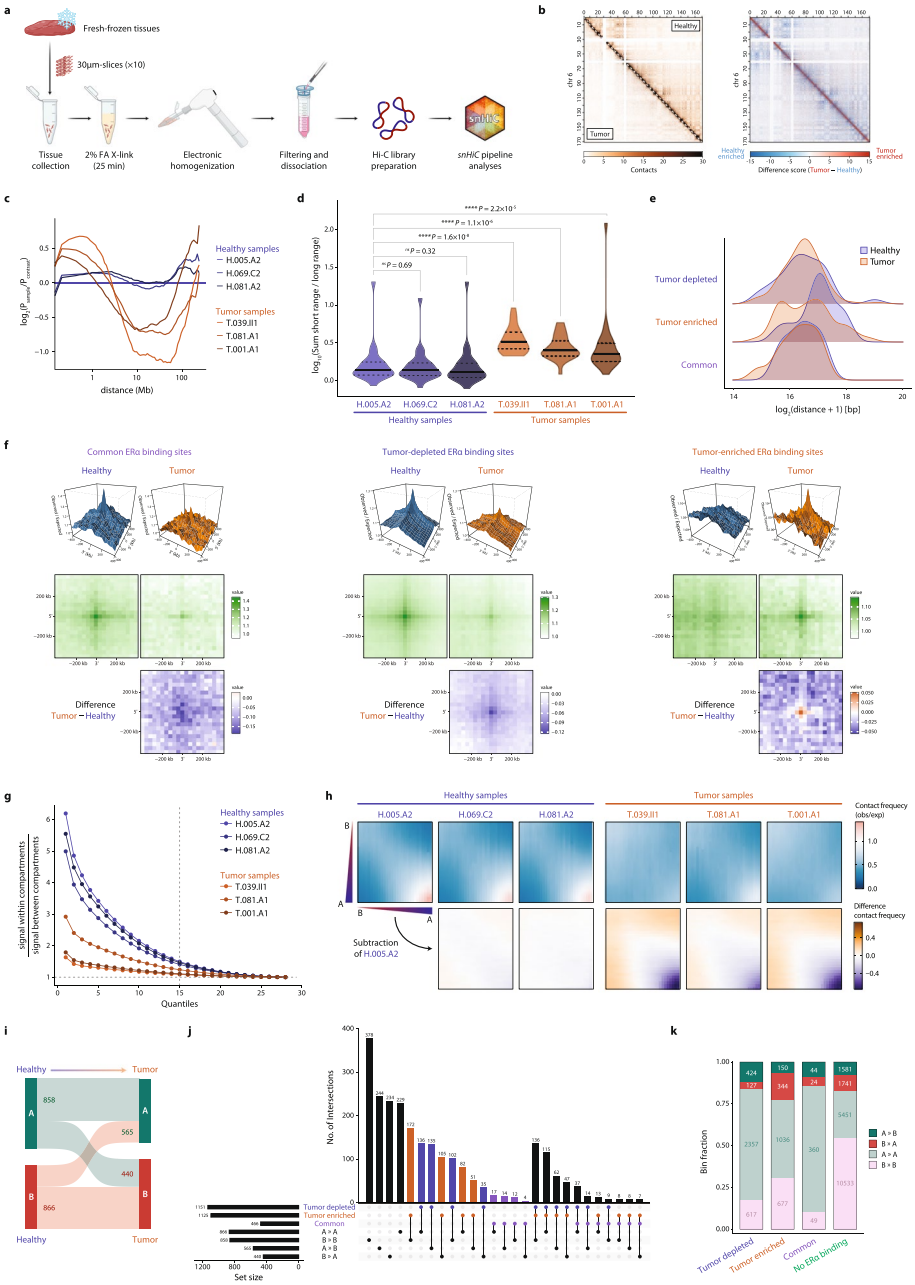

Fig. 2 (See legend on previous page.)

bound in a tissue type-specific manner, we overlapped the compartments with the differential ERα occupied regions upon tumorigenesis (Fig. 2j). We found that compartments switching from A to B class were enriched for tumor-depleted ERα binding sites, relative to tumor-enriched ones (A-to-B depleted/enriched ratio 2.65 (135/51)). In contrast, compartments switching from B-to-A were clearly enriched for tumor-gained ERα binding sites (B-to-A depleted/enriched ratio 0.33 (35/105)). These findings were further confirmed studying the behavior of smaller genomic regions (compartment bins) relative to differentially occupied ERα binding sites, excluding regions that harbor multiple ERα binding site categories (Fisher's exact test: B-to-A depleted-vs-enriched: $P = 1.317$

$\times 10^{-56}$, OR 0.20; A-to-B depleted-vs-enriched: $P = 6.013 \times 10^{-11}$, OR 1.87) (Fig. 2k and Additional file 1: Fig. S4i).

Taken together, our findings highlight a widespread reorganization of the 3D genome upon tumorigenesis that leads to a loss of chromatin compartment polarization. This was accompanied by an extensive class switching, with tumor-enriched ERα binding sites being enriched in B-to-A compartments. Moreover, we identified a gain of relatively shorter-range chromatin interactions upon tumor development, at the expense of longer-range contacts.

Somatic mutations are selectively found at a higher frequency at tumor-enriched ERα binding sites

Recently, pan-cancer whole-genome analysis highlighted the presence of driver mutations in non-coding regulatory elements [25]. To determine whether that might also be the case in endometrial cancer, we re-analyzed whole-genome sequencing (WGS) data of 41 primary endometrial cancer samples from the TCGA Uterine Corpus Endometrial Carcinoma (UCEC) cohort [26] (see " Methods" for details on the case selection). Most of the mutations were located in intronic and intergenic non-coding regions (Fig. 3a). Mutational signatures (Fig. 3b) and variant counts (Fig. 3c) well correlated with the micro-satellite stability (MSS) status of the source samples. These analyses show that most of the micro-satellite instable (MSI) samples are characterized by UV light exposure (SBS7c) and defective DNA mismatch repair (SBS15/44) signatures (Fig. 3b), accompanied by a higher number of mutations per sample (Fig. 3c). On the other hand, micro-satellite stable (MSS) samples, as expected, display a lower number of somatic mutation counts (Fig. 3c) and prevalently mutational signatures associated to defective DNA (polymerase ε exonuclease domain mutations, SBS10b) (Fig. 3b).

To investigate whether somatic mutations in primary tumors are associated with selective ERα activity, we compared the number of somatic mutations that are overlapping with the different categories of ERα binding sites. We found that tumor-enriched binding sites have a higher representation of somatic mutations over the tumor-depleted ones (Fisher's exact $P = 1.143 \times 10^{-5}$, OR 1.89, 95% CI 1.41–2.53; Fig. 3d). Notably, the somatic variants overlapping with the ERα binding sites are carried almost exclusively by hypermutated samples (Additional file 1: Fig. S7a). Interestingly, analyzing the enrichment of genome-wide association study (GWAS) loci associated with endometrial cancer risk at ERα-binding sites [27], we observed that tumor-enriched binding sites are significantly enriched ($P_{adj} = 0.006$) for endometrial cancer risk single-nucleotide polymorphisms (rSNP) in comparison to tumor-depleted regions ($P_{adj} = 0.995$) (Fig. 3e). These data imply that both germline as well as somatic variants are not equally distributed over the genome, but rather show a selective occurrence at tumor-gained ERα sites.

As somatic mutations were enriched at tumor-enriched ERα binding sites, we explored whether the altered ERα binding in tumor samples may be impacted by perturbation and/or acquisition of ESR1 motifs upon tumorigenesis. To test this, we generated a "mutated genome" introducing, in silico, the somatic mutations identified in primary endometrial tumors from the TCGA-UCEC [26] cohort. Subsequently, ESR1 motifs were scored in reference versus our "mutated genome," for all three categories of ERα binding sites (common, tumor-enriched, tumor-depleted), to identify possible

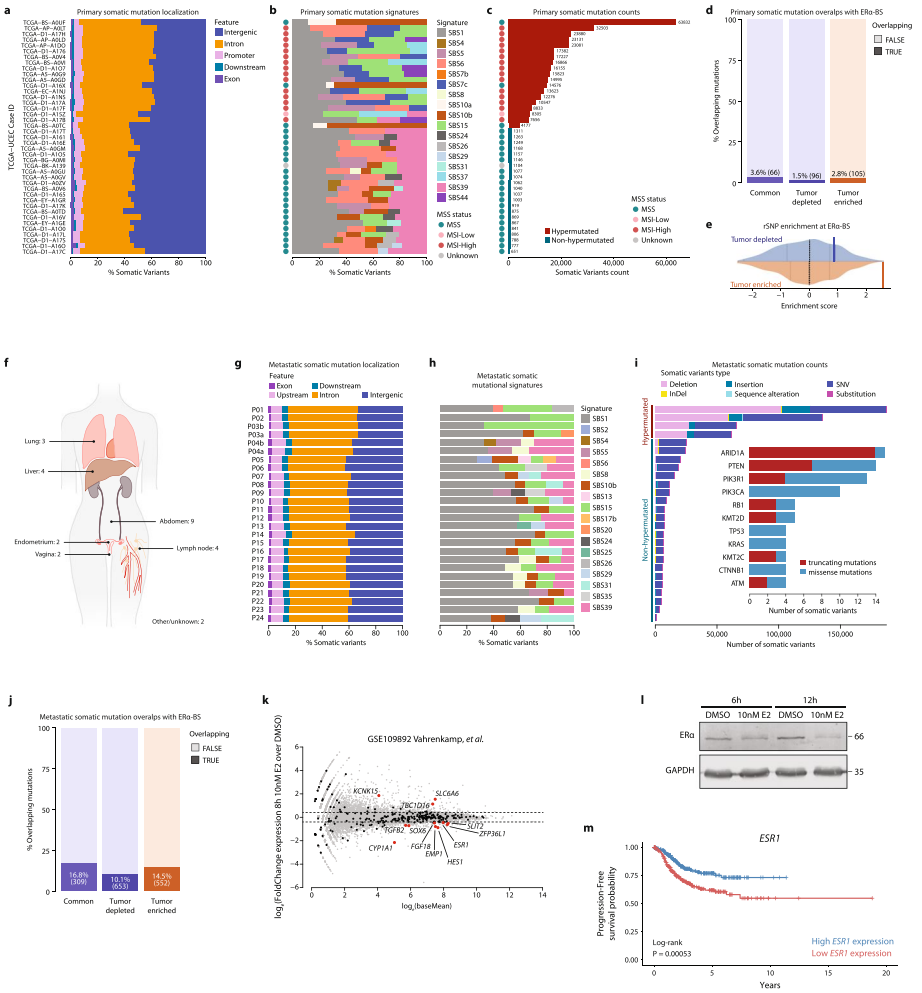

Fig. 3 Tumor-enriched ERα binding sites represent non-coding regions target of somatic mutation in metastatic endometrial cancer tumor. Genomic localization (**a**), mutational signature (**b**), and counts (**c**) distribution of the somatic mutations detected by WGS in the primary endometrial cancer samples form a selection of samples from the TCGA-UCEC cohort ($n = 41$). **d** Stacked bar blot depicting the fraction of ERα peaks overlapping with somatic mutations identified in the primary cohort. **e** Variant set enrichment (VSE) analysis depicting the enrichment of differentially bound ERα sites over endometrial cancer risk loci identified by genome-wide association study (GWAS, $P < 1 \times 10^{-5}$) [27]. Density plot represents distribution for the *Z*-score from matched controls defining the null distribution. Blue (tumor-depleted) or orange (tumor-enriched) vertical lines represent observed enrichments, with tumor-enriched enrichment being statistically significant ($P_{adj} = 0.0060$). Gray vertical lines define 0, 25, 75, and 100 percentiles of the distribution, while dotted black line indicates the median. **f** Scheme of the number metastases and metastatic site, for all metastatic samples used for the WGS analyses. Genomic localization (**g**) and mutational signature (**h**) distribution of the somatic mutations detected by WGS in the metastatic samples described in **f**. **i** In the outer plot, it is shown the number of different types of somatic variants detected in each metastatic sample. In the inner plot, stacked bar plot of the most-frequently protein sequence mutated genes and relative number and type of somatic variants identified in metastatic samples. **j** Stacked bar blot depicting the fraction of ERα peaks overlapping with somatic mutations identified in the metastatic cohort. **k** MA plot of differential expression analyses in Ishikawa cell lines upon 8 h of 10 nM β-estradiol (E2) stimulation. Black dots indicate genes which promoter has been linked, by H3K27ac Hi-C analyses, to a tumor-depleted or tumor-enriched ERα-bound regulatory element bearing somatic mutations in metastatic samples ($n = 311$). Differentially expressed genes (|Fold Change expression| ≥ 1.5 and $P_{adj} < 0.05$) upon estradiol stimulation have been labeled and highlighted by a red dot ($n = 12$). **l** Ishikawa endometrial cancer cells were treated or not for 6 or 12 h with 10 nM estradiol. Whole-cell extracts were analyzed by immunoblotting with antibodies against ERα and GAPDH (loading control). **m** Progression-free Kaplan-Meier curve of endometrial cancer patients (TCGA data) divided into two groups using the median of *ESR1* expression (FPKM) as cutoff

gain-of-function (GOF) or loss-of-function (LOF) effects on ESR1 motifs (Additional file 1: Fig. S8). To define the regions displaying a GOF, we selected regions that do not show any motif ($PWMScore_{ref} = 0$) in the reference, but display a positive score for the "mutated genome" regions ($PWMScore_{mut} > 0$) (Additional file 1: Fig. S8a). Inversely, for LOF regions, we analyzed regions that present ESR1 motifs in the reference ($PWMScore_{ref} > 0$) and that lose this motif upon introduction of the mutations ($PWMScore_{mut} = 0$) (Additional file 1: Fig. S8b). ESR1 motif LOF was observed for a small subset of tumor-depleted regions (22.9%: 200/872, cutoff: $PWMScore_{ref} > 3 \times 10^5$), suggesting that decreased ERα binding at these regions may be invoked by perturbation of its ERE. In contrast, neither GOF nor LOF was observed at the tumor-enriched binding sites.

To investigate whether somatic variant enrichment for particular ERα sites also persisted after endometrial cancer progression, we next analyzed WGS data of 26 fresh frozen biopsy samples, derived from 24 cases of advanced endometrial cancer. Biopsies were collected from tumors that metastasized to lung, liver, and other abdominal regions (Fig. 3f and Additional file 4: Table S3). In these tumors, the vast majority (>96%) of variants were located at non-coding regions, with 85% of the identified variants found to populate intronic and intergenic regions. Considerably, less variants (mean 9%, range 8.1–10.3) were located proximally upstream of a coding gene, potentially representing promoter regions (Fig. 3g). To assess WGS data quality and explore additional genomic features of our samples, we investigated mutational signatures, counts, and frequently mutated genes (Fig. 3h–i). Four samples were hypermutated and displayed 20–67% of variants associated with defective DNA mismatch repair (SBS15). Six samples from tumors pre-treated with carboplatin showed traces of signatures 31 and 35, both associated with platinum treatment [28]. Cumulatively, these results confirm previously described enrichment of tumor-intrinsic [29, 30] and treatment-induced [28] mutational features of endometrial cancer. The most-frequently mutated genes with protein sequence altering mutations included *ARID1A* and various members of the PIK3CA pathway, such as *PTEN*, *PIK3CA*, and *PIK3R1*, recapitulating previous observations for endometrial tumor [29].

As expected, analogously to the primary tumor analyses (Fig. 3a), somatic mutations in metastatic samples were also predominantly occurring in non-coding regions. Next, we integrated these somatic mutation data with tissue-type enriched ERα ChIP-seq data (Fig. 3j), and again found differentially bound ERα regions were significantly enriched for somatic mutations (11.9%, 1221/10,292), relative to the total number of ERα-bound sites from our entire cohort, not found to differ between normal tissue and tumors (8%, 5881/73,312; Fisher's exact $P = 5.64 \times 10^{-36}$, OR 1.54, 95% CI 1.44–1.65). All samples contributed to this enrichment, with a median somatic variant count of 18 (range 1–358) in differentially ERα-bound sites. Moreover, comparing tumor-enriched with tumor-depleted regions, we observed a significant enrichment of somatic variants in the tumor-enriched ERα sites compared to tumor-depleted ones (14.5% vs 10.1%, Fisher's exact $P = 2.98 \times 10^{-11}$, OR 1.51, 95% CI 1.34–1.71, Fig. 3j). In contrast to primary tumors, these observations are independent of the mutational frequency status (Additional file 1: Fig. S7b). In most instances ($n = 1002$), regions overlapped with only one somatic variant, while some ERα-bound regions

were observed to harbor 2 (183 regions), 3 (34 regions), or 4 (2 regions) somatic mutations.

To investigate genes whose expression may be affected by enhancer mutations, we intersected tumor-depleted and tumor-enriched ERα binding sites (FDR $\leq$ 0.01) bearing mutations with H3K27ac HiChIP chromatin looping data in Ishikawa cells (Additional file 5: Table S4), aimed to couple putative enhancers with the transcription start sites (TSS) they control, yielding a total of 331 genes. Out of these, 13 genes were differentially expressed in Ishikawa endometrial cancer cells upon 8 h of estradiol stimulation [31] (Fig. 3k). Interestingly, *ESR1* (encoding ERα) was represented as well, being modestly downregulated at both RNA and protein level (Fig. 3k–l and Additional file 1: Fig. S9c, Additional file 6), and harboring 3 metastatic somatic mutations in two *cis*-regulatory elements upstream of the *ESR1* locus, hereafter referred as "*Enhancer 1*" (P17 and P22; non-hypermutated) and "*Enhancer 2*" (P03b; hypermutated) (Fig. 4a and Additional file 1: Fig. S9b). Of note, high *ESR1* expression is associated with favorable outcome in primary endometrial cancer patients (TCGA [26]) (Fig. 3m). Of note, these two enhancers are specific to the endometrium cancer samples, relative to healthy endometrial tissues and breast cancer specimens, based on ATAC-seq (TCGA [32]) and ERα ChIP-seq signal (our previous work [33]) (Additional file 1: Fig. S9b–c) at the *ESR1* locus.

ESR1 enhancer mutation alters 3D genome contacts and decreases EHMT2/G9a enhancer binding to boost ERα expression

Since *ESR1* expression is strongly associated with favorable outcome in endometrial cancer (Fig. 3h), alterations in its transcriptional regulation may have direct implications on tumor development and progression. Since both *Enhancer 1* and *Enhancer 2*—found mutated in metastatic, but not primary endometrial cancer—showed induction of ERα binding and H3K27ac signals in tumors (Fig. 4a), we investigated whether these loci may undergo alternative regulation in different stages of tumor progression leading to *ESR1* expression deregulation. Of note, in both our RNA-seq data (Additional file 1: Fig. S6a) and the publicly available TNM plot analyses [34] (Additional file 1: Fig. S6b), *ESR1* expression in primary tumors was comparable to the levels found in normal tissue.

Hi-C analyses of endometrial tumors (Fig. 4b) illustrated increased short-range and decreased long-range chromatin interactions at this locus, when compared to healthy endometrial tissue. These data were confirmed with 4C-seq analyses using the *ESR1* promoter as view-poiont (Fig. 4c), showing tumor-specific gained interactions of newly ERα-engaged enhancers with the *ESR1* promoter (Additional file 1: Fig. S4j).

To identify which *ESR1* enhancer mutation had a higher probability to impact *ESR1* gene regulation and potentially drive tumor progression, we employed the *Akita* [35] machine learning tool to predict the effect of the somatic mutations on the 3D genome organization (Fig. 4d). Our analyses predicted the mutation occurring at position chr6:152,002,679 (TTC-to-T, P22) of *Enhancer 1* to affect chromatin architecture surrounding the *ESR1* locus. Based on these results, this specific mutation was selected for further analyses.

To determine whether enhancer mutations can alter the composition of the transcription complex recruitment, we performed DNA affinity purification experiments coupled to quantitative mass spectrometry. To this end, we incubated nuclear extracts from Ishikawa endometrial cancer cells with immobilized DNA-oligonucleotides encompassing

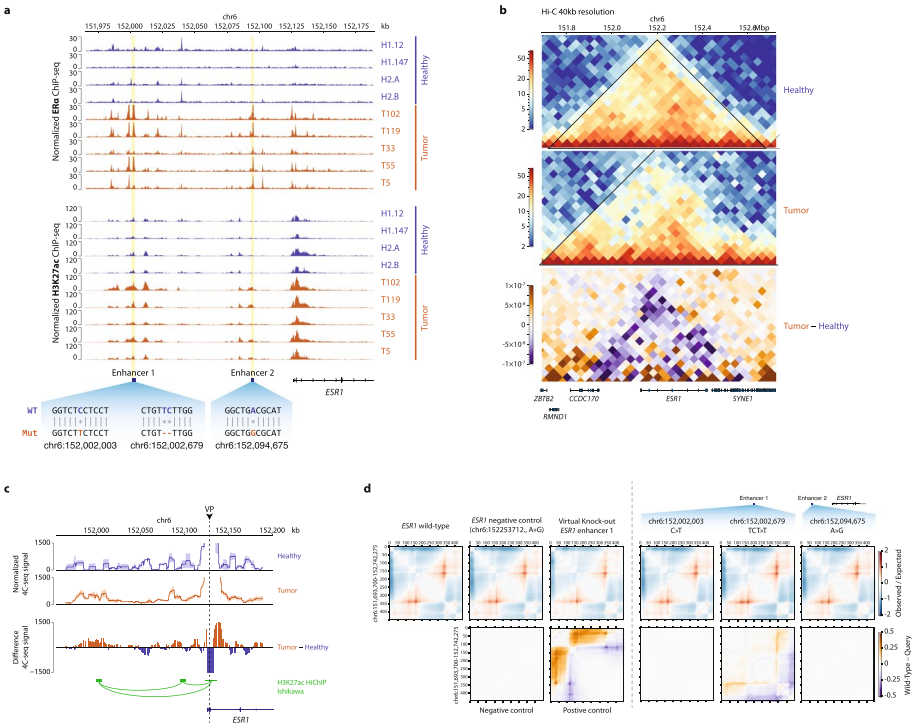

Fig. 4 Short-range chromatin contacts at the *ESR1* locus are stronger in tumors. **a** ChIP-seq genomic tracks for ERα (upper block) and H3K27ac (lower block) in healthy (blue) and tumor (orange) endometrial primary tissues at the *ESR1* locus. *ESR1 Enhancer 1* and *Enhancer 2* are indicated, as well as the mutations found by WGS analyses in metastatic samples. **b** Representation of the averaged 40-kb resolution Hi-C matrix at the *ESR1* locus for the 3 healthy (top) and 3 tumor (middle) tissues or the score difference (bottom) tumor − healthy, where orange indicates higher scores in the tumors while purple higher scores in healthy. The matrices scores are dived by the sum of the matrix. Black lines indicate the topologically associated domains (TADs) identified. **c** 4C-seq genomic tracks at the *ESR1* locus using the *ESR1* TSS as view-point (VP) for 2 healthy (top, blue) and 3 tumor (middle, orange) endometrial tissues, and the average difference of score tumor − healthy (bottom). With green arcs are depicted loops detected by H3K27ac HiChIP in Ishikawa endometrial cancer cells. The ribbon around the 4C-seq signal lines indicates the standard error mean (SEM) among biological replicates. **d** In the top row, observed/expected matrices of sequence-based machine learning prediction in a ±500-kb window surrounding the *ESR1* locus. In order from left to right can be found: wild-type sequence, point mutation in a genomic desert (negative control), deletion of the full *Enhancer 1* sequence (positive control), introduction of SNVs found by WGS analyses of metastatic endometrial cancer samples. On the bottom row is showed the difference of observed/expected score over the wild-type sequence, where orange indicates higher scores in the altered sequence and purple a higher score in the wild-type one

the center of the ERα-bound site at the *Enhancer 1*—containing either the reference or mutant sequence—with a ±25-bp window around it (Fig. 5a–b). Quantitative proteomics identified 13 gained and 11 lost proteins upon mutation of the regulatory element, respectively (Fig. 5b). To identify differentially bound proteins that may directly regulate *ESR1* expression, we used publicly available RNA-seq (TCGA [26]) data of 589 endometrial cancer tissues to correlate the expression of *ESR1* with the expression of genes corresponding to the 24 differentially bound proteins (Fig. 5c). Then, we selected the genes that most-strongly positively (*POLK*, *ZBTB21*, *XPA*, *SIN3A*) or negatively (*GTF2IRD1*, *EHMT2*, *ZNF768*) correlated with *ESR1* expression (Additional file 1: Fig. S10a). For these genes, expression was analyzed in relation to the progression-free survival probability of endometrial cancer patients (TCGA [26]) (Fig. 5d and Additional file 1: Fig.

S10b). Interestingly, our analyses identified only one of the 7 genes analyzed—the lysine methyl-transferase EHMT2 (also known as G9a or KMT1C)—to significantly associate with a poor outcome of endometrial cancer patients (Fig. 5d and Additional file 1: Fig. S10b).

EHMT2/G9a interacts with ERα (Fig. 5e and Additional file 1: Fig. S10c, Additional file 7), based on co-immunoprecipitation experiments. In line with this, ChIP-qPCR revealed that EHMT2/G9a was recruited to both *Enhancer 1* and promoter of the *ESR1* gene in Ishikawa cells, following estradiol treatment (Fig. 5f). As negative control, these experiments were reperformed in the ERα-negative endometrial cancer cell line AN3CA (Additional file 1: Fig. S11a–b). In line with the absence of ERα expression, estradiol treatment did not alter tumor cell growth (Additional file 1: Fig. S11c–d). In absence of ERα, there is little recruitment of EHMT2/G9a to the *ESR1 Enhancer1* and promoter when compared to the negative control region (Additional file 1: Fig. S11e). Furthermore, siRNA-mediated knock-down of EHMT2/G9a in Ishikawa cells resulted in an increase of ERα protein expression (Fig. 5g and Additional file 1: Fig. S10d, Additional file 8), confirming the reciprocal relationship between these proteins. Interestingly, we observed a slight decrease in EHTM2/G9a protein levels upon estradiol stimulation, an effect that was abolished when ERα was targeted for degradation [36] by ICI 182,780 (fulvestrant) treatment (Fig. 5g and Additional file 1: Fig. S10d). As negative control,

(See figure on next page.)

Fig. 5 EHMT2/G9a is a negative regulator of *ESR1* expression in endometrial cancer. **a** Schematic workflow used to perform DNA-oligo protein pull-downs. Biotin-conjugated wild-type or mutated DNA-oligos are immobilized on streptavidin magnetic beads and mixed with Ishikawa nuclear lysates. Captured proteins are then dimethyl labeled and analyzed by mass spectrometry. **b** A DNA oligo with the sequence of the ±25 bp surrounding the ERα binding site in the *ESR1 Enhancer 1*, in wild-type or chr6:152,002,679–TCT-to-T form, was used to perform DNA-oligo protein pull-downs in Ishikawa cells as described in **a**. The scatter plot shows the $\log_2$ ratios of all identified and quantified proteins in both experiments plotted against each other. Proteins significantly enriched at the wild-type sequence are highlighted in red, and proteins significantly enriched at the mutant sequence are highlighted in blue. **c** Gene expression correlation heatmap for all corresponding 24 differentially bound proteins identified in **b** in 589 endometrial cancer patients (TCGA). Dashed lines indicate the separation between positive and negative correlation scores. Genes are ranked by correlation with *ESR1* gene expression. **d** Progression-free Kaplan-Meier curve of endometrial cancer patients (TCGA data) divided into two groups using the median of *EHMT2/G9a* expression (FPKM) as cutoff. **e** EHMT2/G9a and IgG (negative control) ChIP followed by western blot in Ishikawa endometrial cancer cells. Antibodies against EHMT2/G9a and ERα have been used. For EHMT2/G9a, two different exposure images were used for input and IP, as indicated by the vertical dashed bar. **f** EHMT2/G9a ChIP in Ishikawa cells stimulated for 6 h with 10 nM β-estradiol. Bar plot shows percentage of enrichment over the input (% Input) at the *ESR1 Enhancer 1*, *ESR1* promoter, and *CDK12* promoter (negative control) analyzed by quantitative PCR (qPCR). Mean of 2 independent experiments is shown. **g** Ishikawa endometrial cancer cells were stimulated for 72 h with 10 nM β-estradiol in combination or not with 100 nM ICI-182,780 (fulvestrant, negative control). In all conditions, cells where incubated either with a non-targeting (NT) siRNA or with an siRNA against *EHMT2/G9a*. Then, whole-cell extracts were analyzed by immunoblotting with antibodies against ERα, EHMT2/G9a, and GAPDH (loading control). **h** Differential ERα peak centered EHMT2/G9a average ChIP-seq density profiles in Ishikawa cells upon 6 h treatment with 10 nM β-estradiol (purple) or DMSO (blue, control). Ribbon indicates SEM. Paired Wilcoxon test was performed on the average score of the highlighted area for each individual region analyzed; *P* values are indicated. **i** Schematic representation of the *ESR1* gene regulation working model. Upon tumorigenesis, ERα is re-located to an endometrial cancer specific *ESR1* enhancer (Fig. 1f–g). ERα interacts with EHMT2/G9a (**e**) and binds both enhancer and promoter regions of *ESR1*, in a hormone-dependent fashion (**f**). In a subset of metastatic endometrial cancers, a somatic mutation is acquired at the tumor-specific *ESR1* enhancer (Fig. 4a). In vitro analyses probing this mutation revealed a loss of EHMT2/G9a DNA binding capacity (**b**), and EHMT2/G9a knockdown in endometrial cancer cell lines leads to an increase of *ESR1* expression levels (**g**)

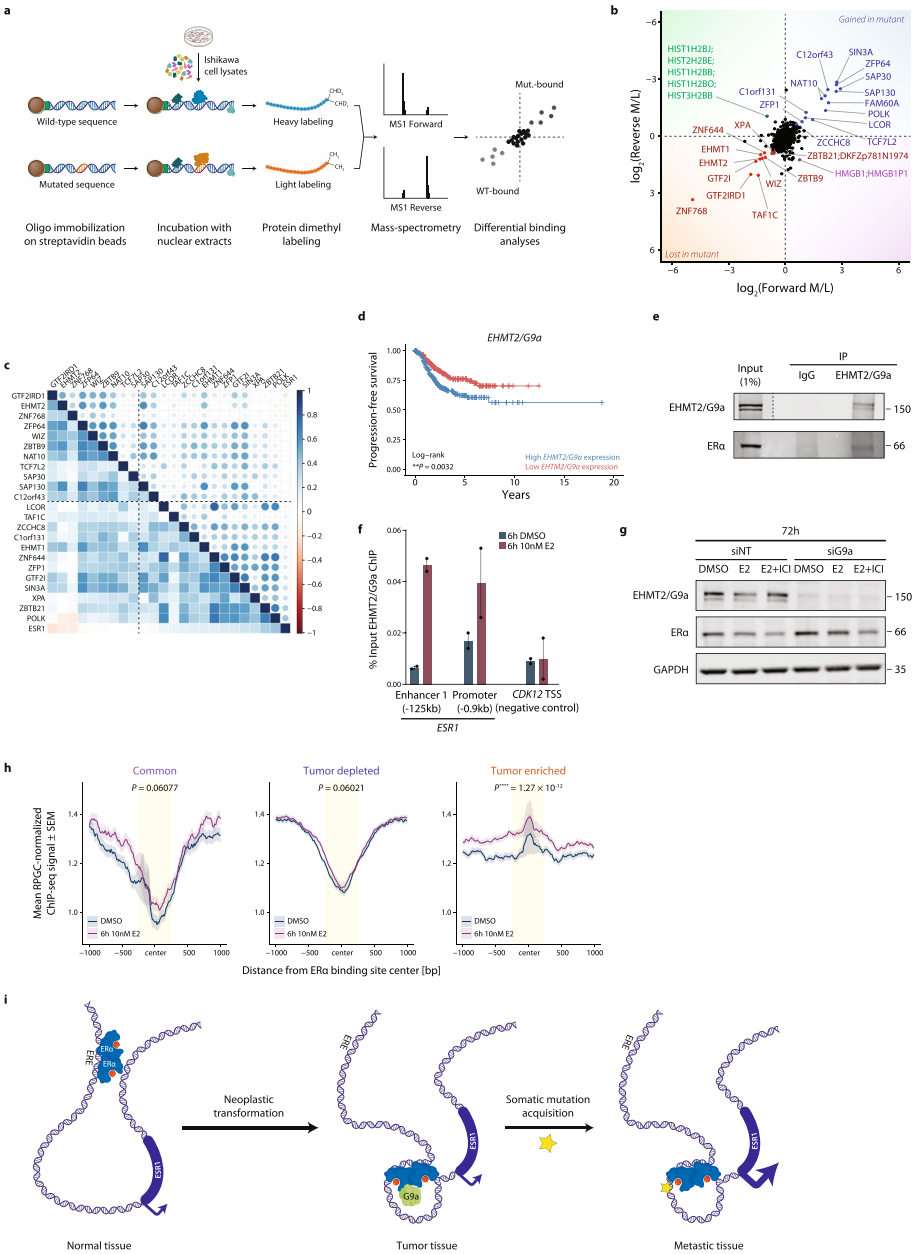

Fig. 5 (See legend on previous page.)

comparable experiments performed in ERα-negative AN3CA cells, which did not show any gain of ERα protein levels upon EHMT2/G9a knock-down (Additional file 1: Fig. S11f–g).

Finally, we wondered whether the ERα-dependent chromatin recruitment of EHMT2/G9a is a specific feature of *ESR1* locus, or rather represents a more general feature of EHMT2/G9a action. To address this question, we studied the EHMT2/G9a ChIP-seq signal around the different categories of ERα binding sites in the presence or absence of estradiol stimulation in Ishikawa cells (Fig. 5h). These analyses revealed an absence of EHMT2/G9a signal at tumor-depleted and common regions, with no differences in its recruitment upon estradiol stimulation. In contrast, EHMT2/G9a co-localizes with ERα

at tumor-enriched binding sites with a significantly increased signal upon estradiol stimulation (Fig. 5h, *right*), suggesting a more general behavior of EHMT2/G9a for tumor-gained ERα sites.

Altogether, these data show that EHMT2/G9a DNA binding is impaired upon mutation of the *ESR1 Enhancer 1* locus, thereby enhancing expression of tumor driver ERα in metastatic endometrial cancer and that its activity is required to repress *ESR1* expression in endometrial cancer in an ERα-dependent manner.

Discussion

High expression levels of *ESR1* (encoding ERα) in endometrial cancer are associated with a favorable outcome (Fig. 3m), since *ESR1*-positive tumors typically proliferate slower as compared to *ESR1*-negative tumors. Nonetheless, for these tumors, ERα plays a critical role in endometrial cancer development and progression [5]; the molecular actions of ERα in this cancer type remain largely unknown. The notion of classical tumor drivers (such as ERα), while being upregulated upon tumorigenesis, can be associated with favorable outcome might feel counterintuitive. However, analogous to breast cancer [37], different subtypes exist for endometrial cancer which differ in their ERα expression status. Patients with ERα-positive tumors typically have a better outcome, as compared to the more aggressive ERα-negative ones [7]. Nonetheless, in these ERα-positive tumors, the receptor is still the oncogenic driver and its activation status drives tumor growth and progression. Therefore, while expression of ERα is considered a favorable prognostic marker, enhanced activity or expression levels of the ERα can drive tumor progression in this subtype.

Through multi-omics analyses (Fig. 1a), we investigated how ERα activity is altered in endometrial cancer and how this feature may be leveraged by tumor cells to progress into the metastatic stage. Cumulatively, our findings allowed us to propose a molecular model describing the dynamic nature of ERα action in endometrial cancer development and progression (Fig. 5i): (i) in healthy endometrial tissue, ERα regulates genes involved in cell cycle and DNA damage response (Additional file 1: Fig. S1 g and Additional file 1: Fig. S3c); (ii) during tumorigenesis, ERα re-localizes to the tumor-specific enhancers (Fig. 1f–g), including the *ESR1* enhancer itself (Fig. 4c). At this enhancer ERα interacts with EHMT2/G9a (Fig. 5e) in estradiol depend manner (Fig. 5f); (iii) during tumor progression, the *ESR1* enhancer acquires a somatic mutation for a subset of tumors (Fig. 4a). In vitro analyses revealed that this mutation impairs EHMT2/G9a DNA binding (Fig. 5b), and perturbation of EHMT2/G9a in Ishikawa endometrial cancer cells enhanced *ESR1* expression (Fig. 5g). Given the low frequency of SNVs at ERα binding sites, many other paths to metastasis may exist, next to the mechanism we propose in this study. As our proposed model is based on integrated datastreams obtained from different models (primary and metastatic patient tissues, cell lines, in vitro experiments), direct causal relations would need to be confirmed in future studies using single model systems, when available. Another limitation of the model is that—due to tissue availability limitations—we had to make use of different cohorts of patients to analyze ChIP-seq and RNA-seq/Hi-C data. While these results are consistent and provide a concordant view between data streams, the use of different patient cohorts limits the complete transferability of our findings.

Our analysis of the ERα cistrome in both normal and neoplastic endometrial tissues revealed limited overlap of chromatin occupancy between the two tissue states. Importantly, despite the healthy and tumor tissues were not matched for the same individual patient, large-scale programmatic changes were consistently observed on chromatin binding of ERα, in comparing the two tissue types. Performing these analyses in matched healthy and tumor samples might lead to the identification of a higher number of significantly differentially bound sites, as noise and inter-sample variation is likely lower. This notion implies that our findings may reflect the strongest effects induced by ERα reprogramming that are to be confirmed in larger future studies. Interestingly, tumor-depleted ERα binding sites did not display differences in H3K27ac, in contrast to the tumor-enriched ones which instead showed a consistent increase in this active histone mark. These results highlighted a specific gain-of-function for newly engaged ERα-bound *cis*-regulatory elements in tumors, while the regulatory elements for which ERα binding is lost remain active through action of other transcription factors. Surprisingly, studying the genomic distribution of the ERα binding sites, we observed an unexpected enrichment of promoter regions for the tumor-depleted binding sites over the tumor-enriched ones. This promoter enrichment is in contrast to prior studies that reported ERα to primarily occupy promoter-distal *cis*-regulatory elements rather than promoters [10, 14, 38, 39]. Importantly, as the molecular biology of this nuclear receptor has been studied mainly using cancer models, the non-tumor context of ERα action remains less well known and may involve a higher degree of promoter involvement.

Our data, combined with previous findings showing that cell-specific ERα sites lack high-affinity estrogen responsive elements (EREs) [11], suggest that tumor-depleted binding sites (enriched for ERα/β motifs) are essential for the cell maintenance in normal tissues. However, during tumorigenesis these sites progressively lose ERα binding in favor of sites enriched in tumors (low-affinity EREs), likely driving cell transformation. Interestingly, at these tumor-enriched binding sites we found an overrepresentation of SOX and FOX transcription factor family motifs. These finding are particularly relevant for the endometrial cancer biology since one of the most recurrently mutated genes in this cancer type is *SOX17* [29], which acts as tumor suppressor [40]. Therefore, we can hypothesize that *SOX17* loss-of-function may leave low-affinity ERα binding sites more accessible for ERα, increasing receptor occupancy at these regions. On the other hand, enrichment of motifs for forkhead pioneer transcription factor class (FOX family) at tumor-enriched ERα binding sites is consistent with the increased H3K27ac signal at these low-affinity ERα binding sites. FOXA1 is known to facilitate ERα binding through its pioneer activity in breast cancer [41], and we previously proposed FOXA1 to serve a comparable role in endometrial cancer [10]. In support of these hypothesis, tumor-enriched ERα binding sites display an enrichment in FOXA1 ChIP-seq peaks detected in Ishikawa endometrial cancer cells (Fig. 1k–l and Additional file 1: Fig. S1j). Furthermore, one can speculate on this mechanism to represent a more general feature of hormone-dependent cancers. Indeed, comparison between healthy prostate epithelium and prostate adenocarcinoma—being strongly dependent on androgen receptor (AR) action—showed tumor-gained AR binding sites to be selectively enriched for FOXA1 motifs, as opposed to AR sites found exclusively in healthy tissues [16].

Enhancers interact with promoter regions to regulate gene expression, facilitated through spatial rearrangement of the chromatin structure [21]. Performing Hi-C analyses in healthy and tumor tissues, we observed a global enrichment of short-range (<2 Mb) chromatin interactions, at the expense of long-range (>2 Mb) contacts. Moreover, chromatin loop distance in tumor tissues is reduced at tumor-enriched ERα binding sites when compared to normal samples, while loop distance at shared and tumor-depleted sites remains unchanged. This corroborates with the observed newly acquired H3K27ac signal at tumor-enriched ERα binding sites in tumor samples in contrast to common and tumor-depleted sites, where this active enhancer/promoter histone mark is persistent in both tissue types. Combined, these findings imply a gain of enhancer activity at tumor-enriched binding sites that may lead to an increase in enhancer-promoter proximal contacts, explaining the loop distance shortening.

Next to the short-/long-range interaction alterations between heathy and tumor tissues, we also observed a compartment depolarization in tumor samples. Chromatin compartment destabilization is observed in many tumor types [42] (e.g., colorectal cancer [43] and glioblastoma [44]). Interestingly, as tumor-enriched ERα binding sites were enriched at B-to-A compartment switched sub-regions, these data suggest that alterations in 3D genome organization directly contribute to enhancer plasticity in tumorigenesis, concordant with increased H3K27ac signal at these sites.

As the vast majority of mutations are found at non-coding regions of the genome, we investigated whether tumor-enriched ERα sites could drive tumor progression through acquisition of mutations. Through integration of somatic mutation data with RNA-seq in Ishikawa cells, we concluded that most genes with proximal ERα-bound enhancer mutations are not under direct ERα control. This is in agreement with our previous observation in prostate cancer, in which only a minor fraction of SNVs at regulatory elements functionally affected their activity [17].

We also observed a depletion of somatic mutations in tumor-depleted ERα binding sites, relative to the common and tumor-enriched ERα sites (Fig. 3d, j). We could speculate that two mechanisms may underlie this observation. The first is based on previous reports showing that, since DNA-bound proteins prevent efficient DNA repair [45], somatic mutations preferentially accumulate at active TF-binding sites. On the other hand, accumulation of functional mutations may occur at tumor-enriched sites by other mechanisms. In our previous work in prostate cancer [17], we analogously observed that tumor-gained AR binding sites were more often somatically mutated, but were also more-often positive for cancer risk SNPs. These results put our current study in perspective, suggesting that the observed higher frequency of SNVs in tumor-gained regulatory elements for hormone receptors may be a feature that is common to hormone-dependent cancers.

While the interaction between ERα and EHMT2/G9a has already been reported in breast [46, 47], EHMT2/G9a is considered to be an ERα co-activator in that cellular context. However, studies in erythroleukemia [48, 49] describe a dual role of EHMT2/G9a as both gene activator and repressor depending on its interaction with mediator rather than JARID1, respectively. As we observed ERα expression levels being upregulated following knockdown of EHMT2/G9a (Fig. 5g), an ERα co-repressive role for EHMT2/G9a in the endometrial cancer cell context is supported by our data. Due to the lack of

other ERα-positive endometrial cancer cell line availability [50], only Ishikawa cells were used in this study. Therefore, despite experiments carried out in AN3CA ERα-negative cells support our findings, further investigation would be required to further substantiate what determines the repressive transcriptional effect of EHMT2/G9a in the interplay with ERα in this tumor type. Moreover, despite many data streams were generated in tumor tissues, EHMT2/G9a functional experiments performed in Ishikawa cells—deriving from endometrial epithelial cells—places our conclusion in an intrinsically tumor cell-centric context that ignores any potential influence of the microenvironment (e.g., stromal and immunity cells). Further studies would be required to determine whether the endometrial cancer stromal compartment may influence ERα regulation in the tumor itself.

In this study, we shed light on the molecular mechanisms underlying endometrial cancer development and progression, connecting alterations in 3D genome organization with epigenetic plasticity and non-coding somatic mutations, using *ESR1* enhancer mutation as an example. This study may serve as blueprint for studies in other hormone-driven cancer types, in which such complex connections are yet to be identified.

Conclusion

Collectively, this study revealed that endometrial tumorigenesis involves an extensive rewiring of the regulatory elements occupied by ERα, which are specifically gained in active epigenetic marks in cancer samples. These tumor-enriched sites are enriched for non-coding somatic mutations in both primary tumors and metastatic samples. One of these mutations occurs at an endometrial cancer-selective *ESR1* enhancer which is regulated though the recruitment of G9a/EHTM2 in an ERα-dependent manner.

Methods

Human tissue collection and quality assessment

Fresh-frozen healthy and tumor samples were obtained from post-operative tissue at the Netherlands Cancer Institute (Amsterdam, the Netherlands) from patients who did not receive neoadjuvant endocrine treatment. Tumor content was assessed by hematoxylin and eosin (H&E) staining on slides taken throughout the tissue sample. For tumors, only samples displaying a tumor percentage greater or equal to 70% were deemed eligible for further analysis. Of note, non-cancerous healthy tissues were obtained from patients whose uterus was removed because of (a) surgery for cervical carcinoma or ovarian cancer and (b) endometrial cancer diagnoses, but the area that was cryo-preserved was devoid of tumor cells.

The study was approved by the institutional review board of the Netherlands Cancer Institute, written informed consent was signed by all participants enrolled in the study, and all research was carried out in accordance with relevant guidelines and regulations.

Cell culture and chemicals

Ishikawa (Merck Sigma Aldrich) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) and DMEM Mixture F-12 (DMEM/F-12, Gibco), respectively, supplemented with 10% fetal bovine serum (FBS-12 A, Capricorn Scientific) and penicillin/streptomycin (100 µg/mL, Gibco). AN3CA endometrial cancer cells (ATCC) were

cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS-12 A, Capricorn Scientific) and penicillin/streptomycin (100 µg/mL, Gibco). Cell lines were subjected to regular *Mycoplasma* testing and underwent authentication by short tandem repeat profiling (Eurofins Genomics).

For hormone stimulation, cells were pre-cultured for 3 days in phenol red-free DMEM (Gibco, Ishikawa) or DMEM/F-12 (Gibco, AN3CA) supplied with 5% dextran-coated charcoal (DCC) stripped FBS, 2 mM L-glutamine (Gibco), and penicillin/streptomycin (100 µg/mL, Gibco), then stimulated with 10 nM DMSO-solubilized 17β-estradiol (MedChemExpress, #HY-B0141) for the indicated amount of time. Inhibition of ERα activity was performed by treatment of the cells with 100 nM DMSO-solubilized fulvestrant/ICI-182,780 (MedChemExpress, #HY-13636).

Cell proliferation analyses

Cells were plated in a 96-well plate at a density of 10,000 cells/well. Cells were treated with 100 nM DMSO-solubilized fulvestrant/ICI-182,780 (MedChemExpress, #HY-13636). Cells were imaged every 4 h by using an IncuCyte ZOOM Live-Cell Analysis System (Essen BioScience, Sartorious), and cell confluency percentage was calculated using the IncucyteZoom (v2018 A) software. Normalization, by subtraction of the first point (T_0), and visualization were performed using *Rseb* [51] (v0.3.2) R-package (https://github.com/sebastian-gregoricchio/Rseb).

ChIP-qPCR and ChIP-seq library preparation

Chromatin immunoprecipitation of ERα (SantaCruz #sc-543, 5 µg/IP) and H3K27ac (ActiveMotif #39133, 5 µg/IP) were performed as previously described [33] employing the combination of DSG and formaldehyde to perform the crosslinking. The same protocol, but using 1% formaldehyde for the crosslinking, has been used to perform EHMT2/G9a (Abcam #ab133482, 5 µg/IP) and Normal-IgG (Merck Millipore #12–370, same amount than IP) ChIPs in Ishikawa cells.

Immunoprecipitated DNA is compared to input sample by the comparative C_t method, and signal enrichment is expressed as percentage of input (% Input). Real-time quantitative PCR (RT-qPCR) was performed on the QuantStudio™ 5 Real-Time PCR System 384-well (Applied Biosystems, #A28140) using SenSMix™ SYBR® No-ROX buffer (Meridian Bioscience). The following genomic locations have been tested: *ESR1_Enhancer1*−125kb (Fw: 5′-TGGTAGGTGCTCAGGAGATAA-3′, Rv: 5′-CAGCGACTCGAACAGGATTT-3′), *ESR1*_promoter-0.9kb (Fw: 5′-CCACTCCTGGCATTGTGATTA-3′, Rv: 5′-CAGGACACATGACACCCAAT-3′), *CDK12*_TSS (Fw: 5′-GGACCTGATCTCGCGTTGTT-3′, Rv: 5′-TAGCCTCTCGCGATGTTTCG-3′).

ChIP-seq data processing, motif, and Gene Ontology enrichment analyses

Reads were aligned to the human genome build GRCh37 using *BWA* 0.5.9-r26-dev [52]. Reads with a mapping quality (MAPQ) < 20 were removed from further analysis, and duplicates were marked using GATK markDuplicates [53]. Enrichment over input control was determined using both *MACS2* [54] ($q < 0.01$) and *DFilter* [55]. Only peaks identified by both methods, and not overlapping with the ENCODE GRCh37/Hg19 blacklisted regions [56], were retained. ERα consensus peaks for a given condition

(healthy or normal tissues) have been defined as peaks found in at least 75% of the samples.

Differentially bound regions between tumor and healthy samples were identified using the R-package *DiffBind* [19] (https://bioconductor.org/packages/release/bioc/html/DiffBind.html) with *edgeR* [57] mode at an FDR < 0.05. RPGC-normalized (Reads Per Genomic Content-1× coverage: (mapped reads × fragment length)/effective genome size; *bamCoverage* [58]) ChIP-seq signal at peaks was visualized using *deepTools* [58], while genomic tracks and average density profiles were generated with *Rseb* [51] (v0.3.2) R-package (https://github.com/sebastian-gregoricchio/Rseb) or *pyGenomeTracks* [59, 60].

Genomic feature annotations and distance to TSS were performed using *ChIPseeker* [61] (promoter: −2 kb:TSS: +1 kb, flanking distance: 2 kb), while for motif enrichment in peak regions we used *AME* v5.5.05 [62] algorithm from the *MEME* suite [63] (v.5.5.5) using the JASPAR database [64, 65] as reference. Visualization of the enrichment motifs was made using *ggwordcloud* R-package using the *AME* computed *E*-value.

Area proportional Venn diagrams were created using the *Vennerable* (https://github.com/js229/Vennerable) R-package.

Analyses for Gene Ontology biological process (GO-BP) enrichments for tumor-depleted ERα promoter-bound genes were performed using DAVID [66] (v6.8) and employing tumor-enriched ERα promoter-bound genes as background data set. Unsupervised clustering permutation test was performed using *ConsensusClusterPlus* [67] (v1.54.0) on the *DiffBind* normalized counts using the following parameters: reps = 100 (number of permutations), pItem = 0.8 (80% of sample shuffling), pFeature = 1 (100% of features to sample), clusterAlg = "hc" (hclust, hierarchical clustering), distance = "spearman." To determine the ESR1/ERα motif strength at each individual ERα binding site, fasta-formatted sequences of the peaks were obtained using BED2 FASTA from the *MEME* suite [63] (v.5.5.5). Then, *PWMScore*, from *PWMTools* [68], was used to compute the ESR1/ERα motif (JASPAR id: MA0112.3) sum occupancy score for each site. *GIGGLE* [20] analyses were performed using the Cistrome Data Browser Toolkit (http://dbtoolkit.cistrome.org/).

Hi-C library preparation and data processing

Flash-frozen primary tissues have been processed as described in Fig. 2a. Briefly, 10 × 30 µm slices are cross-linked (1 mL of 2% methanol-free formaldehyde in 1× DPBS for 25 min), washed, electronically homogenized and filtered using a 75-µm cell strainer, and collected in a 1.5-mL microcentrifuge loBind tube. Then, Hi-C single-index library preparation was performed as previously described [69] using MboI (New England Biolabs) restriction enzyme.

Quality and quantification of the Hi-C libraries was assessed using the 2100 Bioanalyzer (Agilent, DNA 7500 kit). An equimolar pool of the different samples was sequenced on the Illumina NextSeq 550 System in a 75-bp paired-end setup. De-multiplexed fastq data were analyzed at 2-kb, 10-kb, 40-kb, 100-kb, and 500-kb resolution using the *snHiC* [24] (v0.1.1) pipeline (https://github.com/sebastian-gregoricchio/snHiC) using default parameters and the Hg19/GRCh37 genome assembly. This pipeline relies on *HiCExplorer* [70, 71] (v3.7.2) for the matrix generation, normalization and correction, TAD

and loop detection, and long/short contacts ratio and quality controls. Compartments are identified by *dcHiC* [72] (v2.1). Downstream analyses have been performed using *GENOVA* [73] (v1.0.1), which was used to plot Hi-C contact matrices heatmaps, compute relative contact probabilities, and perform paired-end spatial chromatin analyses (*PE-SCAn* [23]) at ERα binding sites.

To test for chromosomal translocations, we used the *HiNT-TL* function from the *HiNT* tool [74] using a *P* value cutoff of 0.001. Genome-wide normalized and corrected Hi-C heatmaps have been plotted using *hicPlotMatrix* function from *HiCExplorer* [70, 71] (v3.7.2).

HiCExplorer [70, 71] (v3.7.2) *hicCompartmentalization* was used to quantify the compartment polarization. For the compartment switching computation, A and B compartment locations in the reference condition were obtained merging adjacent bins displaying positive or negative, respectively, compartmentalization scores. Then, average compartmentalization score in the two conditions was computed for each reference compartment. A compartment in the reference condition was defined as "switching" when the average compartmentalization score in the reference condition switched sign in the test condition. Compartment-ChIP peaks overlaps and relative upset plot have been computed using *intervene* [75], while Hi-C tracks were generated by *pyGenomeTracks* [59, 60].

Whole-genome sequencing (WGS) analyses

Primary samples

From the full Uterine Corpus Endometrial Carcinoma (UCEC) cohort available in the TCGA [26] database, 41 samples have been selected in order to match the following criteria: gender = "FEMALE," histological_type = "Endometrioid endometrial adenocarcinoma," history_of_neoadjuvant_treatment = "No," horm_ther = "No, I have never taken menopausal hormone therapy.," menopause_status = "Post (prior bilateral ovariectomy OR >12 mo since LMP with no prior hysterectomy)," prior_tamoxifen_administered_usage_category = "Never Used," radiation_therapy = "NO," sample_type = "Primary Tumor," targeted_molecular_therapy = "NO," tumor_tissue_site = "Endometrial," *ESR1* TPM-normalized gene counts ≥ 5. Somatic mutation calling was performed on base-quality recalibrated aligned whole-genome sequencing data available under restricted access on the TCGA portal (data access authorization, project ID: 36269), using blood-derived matched normal samples as germline reference, by *GATK* (v 4.3.0.0) *Mutect2* [53]. Only mutations passing the GATK *FilterMutectCalls* filter and displaying a sequencing depth greater than 20 (DP > 0) were retained.

Metastatic samples

Sampling, sequencing, and initial computational steps have been previously described extensively by Priestley et al. [76] and performed at the central sequencing facility at the Hartwig Medical Foundation.

Frequently mutated genes were determined by overlapping genes with somatic mutations affecting protein coding regions in more than one sample in our dataset, with the mutated genes described in the endometrial carcinoma publication from TCGA [26]

and the OncoKB Cancer Gene List [77]. Overlap between somatic variants and ChIP-seq regions was determined using *bedtools intersect* (v2.25.0) [78].

For both datasets, mutational signatures were determined using the *deconstructSigs* [79] R-package, using the COSMIC mutational signatures v3 database [80]. Genomic feature annotation was performed using *ChIPseeker* [61] (promoter: −2 kb:TSS: +1 kb, flanking distance: 2 kb).

To score the ESR1 motif upon in silico genome mutagenesis (Additional file 1: Fig. S8), we used *bcftools merge* to merge all the VCFs files containing the mutations identified by *Mutect2* in the TCGA-UCEC primary tumors. We then generated a "mutated" genome using *GATK FastaAlternateReferenceMaker*. For both the reference genome and the "mutated genome", we used PWMScore to score the ESR1 motif in reference or mutated conditions for the 3 categories of ERα peaks (tumor-enriched, tumor-depleted, common).

Risk single-nucleotide polymorphism (rSNP) enrichment analyses

We assessed the enrichment of endometrial cancer germline risk variants among differentially bound ERα sites using the *VSE* [81] (v0.99) R-package. Twenty-seven endometrial cancer risk loci (lead variant $P < 1 \times 10^{-5}$) were identified from a genome-wide association study of endometrial cancer [27]. Variants in strong LD ($r^2 > 0.8$) with the lead SNP at each locus were selected using 1000 Genomes Project phase 3 [82] to generate an associated variant set (AVS). A null-distribution was built on the basis of 500 matched random variant sets. We assessed enrichment within tumor-enriched and tumor-depleted ERα binding sites, with and without including a 500-bp window surrounding each binding site. A Bonferroni-corrected P value < 0.05 (adjusting for four tests) was considered statistically significant.

Endometrial human tissue derived RNA-seq library preparation and differential expression analyses

RNA from ~30 mg of endometrial healthy or tumor tissue was extracted using the RNeasy mini kit (Qiagen) following manufacturer instructions. Quality of the RNA extraction was assessed using the 2100 Bioanalyzer (Agilent, RNA 6000 Nano Kit) and selecting samples with a RNA integrity number (RIN) above 9. PolyA+ stranded RNA library was prepared using the Illumina Stranded mRNA Prep kit and quality was assessed by using the 2100 Bioanalyzer (Agilent, DNA 7500 kit). RNA-seq libraries have been pooled equimolarly and sequenced using NovaSeq6000 (Illumina) sequencer with a 51-bp paired-end reads setup. Fastq files have been demultiplexed by *Cutadapt* [83] and mapped on Hg38/GRCh38 genome assembly using *HISAT2* [84] (v2.1.0) using the following parameters: --wrapper basic-0 --min-intronlen 20 --max-intronlen 500000 --rna-strandness FR -k 5 --minins 0 --maxins 500 --fr --new-summary. *HISAT2* [84] (v2.1.0) was used to generate raw gene counts; gene counts normalization and differential expression analyses were performed using *DESeq2* [85] (v1.30.1). Differentially expressed genes were defined by |Fold Change expression| ≥ 2 and $P_{adj} < 0.05$. Data were visualized using *Rseb* [51] (v0.3.2) (https://github.com/sebastian-gregoricchio/Rseb).

Gene set enrichment analyses (GSEA)

Differentially expressed genes were ranked by the $\log_2$(Fold Change expression) calculated using *DESeq2* [85] (v1.30.1) and used to perform GSEA analyses by *clusterProfiler* [86] (v3.18.1) on the "Hallmarks" (H) datasets retrieved through *msigdbr* (v7.5.1). Visualization of the results was performed using *Rseb* [51] (v0.3.2) (https://github.com/sebastian-gregoricchio/Rseb).

Public RNA-seq and differential gene expression analyses on Ishikawa cells

Ishikawa gene expression data was obtained from GEO (accession number: GSE109892 [31]). Differential gene expression between estradiol and DMSO treated cells was computed using *DESeq2* [85] (v1.30.1). Differentially expressed genes were defined by |Fold Change expression| ≥ 1.5 and $P_{adj} < 0.05$.

Circularized chromosome conformation capture sequencing (4C-seq)

For each sample, 10–15 30-µM-thick slices of flash frozen tissue were cross-linked under rotation at room temperature for 20 min in 5 mL of 1× PBS supplemented with 2% FBS and 2% methanol-free formaldehyde. The cross-linked slices have been washed twice in 10 mL of ice-cold 1× PBS before to be resuspended in 0.5–1 mL of ice-cold 1× PBS in 1.5-mL DNA-loBind microcentrifuge tubes. Tissues have been electronically homogenized and lysates filtered using a 75-µm cell strainer. Then, experiments of 4C-seq were performed as previously described [87]. Briefly, nuclei were isolated and permeabilized to allow digestion of the chromatin by the primary RE (DpnII, New England Biolabs). Upon dilution, chromatin fragments have been ligated and then de-crosslinked. DNA was purified and digested with the secondary RE (NlaIII, New England Biolabs) and circularized by ligation. PCR amplification of the re-purified circular fragments have been performed using view-point specific primers (reading primer: 5′-tacacgacgctcttccgatctAACTCGATTTGGAGCGATC-3′; non-reading primer: 5′-actggagttcagacgtgtgctcttccgatctCTGGGACTGCACTTGCTC-3′) using the Expand Long Template PCR System (Roche). Amplicons were purified with AMPure XP beads (Beckman Coulter) in a 0.8× ratio and then amplified using standard indexed Illumina primers as previously described [87] using the Expand Long Template PCR System (Roche). Second-round PCR products were purified with PCR purification columns (Qiagen) and quantified by 2100 Bioanalyzer (Agilent, DNA 7500 kit).

4C libraries have been pooled equimolarly and sequenced using Illumina MiSeq with a 75-bp single-end reads setup. Fastq files have been demultiplexed by *Cutadapt* [83] and mapped on Hg19/GRCh37 genome assembly and signal normalized by *pipe4C* [87] v1.1 R-package in "cis" mode and with default parameters. The signal of different technical replicates was averaged to obtain a unique signal for each sample. The, each tissue sample has been considered as biological replicate. Downstream analyses of 4C-seq data have been performed in an R v4.0.3 environment by using *get.single.base.score.bw* and *genomic.track* functions from *Rseb* [51] (v0.3.2) (https://github.com/sebastian-gregoricchio/Rseb) R-package in combination with *ggplot2* (v3.3.5) and *ggforce* (v0.3.3) R-packages.

Protein extraction and immunoblotting

For immunoprecipitated proteins, elution buffer (1% SDS, 0.1 M $NaHCO_3$) and Laemmli buffer, supplied with 100 mM DTT (dithiothreitol, Merck Sigma Aldrich), were added to input samples and beads, and then boiled at 95 °C for 30 min. For whole-cell extracts, cells have been lysed using 2× Laemmli buffer supplied with 100 mM DTT (dithiothreitol, Merck Sigma Aldrich), boiled at 95 °C for 10 min and then DNA was sheared by sonication (Bioruptor® Pico, Diagenode; 4 cycles of 30 s ON + 90 s OFF). In both cases, proteins were then resolved by SDS-PAGE and transferred to a 0.22-µm nitrocellulose membrane (BioRad). Membranes were incubated for 2 h with blocking solution (BS: 1× PBS, 0.1% Tween20, 5% non-fat milk powder) and then incubated with primary antibody against ERα (ThermoFisher #MA5-14104, 1:1000 in BS), EHMT2/G9a (Abcam #ab133482, 1:1000 in BS), or GAPDH (ThermoFisher #PA1-987, 1:5000 in BS) overnight at +4 °C. After washing with washing solution (1× PBS, 0.1% Tween20), membranes were incubated with secondary antibodies donkey-α-mouse IRDye® 680RD (926–68073, LI-COR Biosciences, 1:10,000 in BS), donkey-α-mouse IRDye® 800 CW (926–32212, LI-COR Biosciences, 1:10,000 in BS), donkey-α-rabbit IRDye® 800 CW (926–32213, LI-COR Biosciences, 1:10,000 in BS), or donkey-α-rabbit IRDye® 680RD (926–68073, LI-COR Biosciences, 1:10,000 in BS) for 1 h. Membranes were washed again with washing solution, scanned and analyzed using an Odyssey® CLx Imaging System (LI-COR Biosciences) and *ImageStudio™ Lite* v.5.2.5 software (LI-COR Biosciences).

Sequence-based machine learning 3D genome folding predictions

The 3D genome organization perturbation predictions have been performed using a window of 2^{20} bp around the *ESR1* locus employing the convolutional neural networks (CNN) model from *Akita* [35] tool. We used the pre-existing model based on Hi-C data originated from human foreskin fibroblast (HFF). Heatmaps have been generated using *matplotlib.pyplot.matshow*.

DNA-oligo protein pull-down

Ishikawa cells were harvested and washed twice with ice-cold PBS and nuclear extracts were prepared as described previously [88]. Briefly, cells are washed with 1× PBS and trypsinized. Trypsin is neutralized by adding the appropriate SILAC medium, and "light"/"heavy" cells are collected separately at 4 °C. Cells are washed in 1× ice-cold PBS and resuspended in five volumes of ice-cold buffer A (10 mM KCl, 20 mM Hepes KOH pH 7.9, 1.5 mM $MgCl_2$). Cells are then pelleted and resuspended in two volumes of buffer A supplemented with protease inhibitors and 0.15% Igepal NP40 (v/v) before to be dounce homogenized. Nuclei are then collected and resuspended in two volumes of buffer C (420 mM NaCl, 20 mM Hepes KOH pH 7.9, 20% glycerol (v/v), 2 mM $MgCl_2$, 0.2 mM EDTA, and 0.1% Igepal CA630 (NP40; Sigma-Aldrich, I8896), EDTA-free complete protease inhibitors (Roche), and 0.5 mM DTT) and incubated for 1 h. The suspension is centrifuged, and the supernatant containing the nuclear extracts is collected.

The ~50-bp oligonucleotide probes encompassing the SNP were ordered with the forward strand containing a 5′-biotin moiety (Integrated DNA Technologies) (Fw_WT: 5′-/5Biosg/ACTGTGGAAACTGGAAGCTGTTCTTGGACTATTTCGCAACACTT

TTCTCC-3′; Rv_WT: 5′-GGAGAAAAGTGTTGCGAAATAGTCCAAGAACAG CTTCCAGTTTCCACAGT-3′; Fw_SNP: 5′-/5Biosg/ACTGTGGAAACTGGAAGC TGTTTGGACTATTTCGCAACACTTTTCTCCTC-3′, Rv_SNP: 5′-GAGGAGAAA AGTGTTGCGAAATAGTCCAAACAGCTTCCAGTTTCCACAGT-3′; mutated site is underlined). DNA affinity purifications, on-bead trypsin digestion and dimethyl labeling were performed as described [89]. Matching light and medium labeled samples were then combined and analyzed using a gradient from 7 to 30% buffer B in buffer A over 44 min, followed by a further increase to 95% in the next 16 min at flow rate of 250 nL/min using an Easy-nLC 1000 (Thermo Fisher Scientific) coupled online to an Orbitrap Exploris 480 (Thermo Fisher Scientific). MS1 spectra were acquired at 120,000 resolution with a scan range from 350 to 1300 m/z, normalized AGC target of 300% and maximum injection time of 20 ms. The top 20 most intense ions with a charge state 2–6 from each MS1 scan were selected for fragmentation by HCD. MS2 resolution was set at 15,000 with a normalized AGC target of 75%. Raw MS spectra were analyzed using MaxQuant software (version 1.6.0.1) with standard settings [89, 90]. Data was searched against the human UniProt database (downloaded 2017) using the integrated search engine. N-terminal and lysine modification for dimethyl labeling was specified under "labels." Carbamidomethylation was specified as a fixed modification. N-terminal acetylation and methionine oxidation were selected as variable modifications.

Transient small interfering RNA (siRNA) cell transfection

Transient transfections of Ishikawa and AN3CA cell lines was performed according to the manufacturer's instructions using Lipofectamine™ RNAiMAX (Invitrogen) for siRNA knockdown experiments. The ON-TARGETplus™ siRNA SMARTpool targeting human EHMT2/G9a (L-006937-00-00050) and the siGENOME™ Non-Targeting control siRNA #5 (D-001210-05-20) were purchased from Dharmacon.

Statistical analysis

All statistical analyses were performed in R versions 3.4.4, 3.5.1, or 4.0.2 (R Core Team 2020, https://www.R-project.org). Enrichment of Ishikawa peaks and somatic mutations in differentially bound regions was calculated using Fisher's exact test and, where appropriate, corrected for multiple testing using the FDR.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03596-5.

Additional file 1: Supplementary Figures S1–S11, including uncropped blot images

Additional file 2: Table S1 Clinicopathological features of patient samples and omics data streams generated. Description of the collection site, histological status, site of collection, age at diagnosis, tumor percentage, grade, and FIGO classification for each patient sample. The last column indicates the omics data generated

Additional file 3: Table S2 List of publicly available ChIP-seq data sets in Ishikawa cells. List of transcription and epigenetic factors ChIP-seq data in Ishikawa cells and corresponding ENCODE accession number

Additional file 4: Table S3 Metadata relative to metastatic endometrial cancer samples. In this tables are reported the fraction of tumor cells, year of birth, biopsy site, and eventual pre-treatments of the 24 metastatic endometrial cancer samples

Additional file 5: Table S4 List of ERα binding sites and H3K27ac HiChIP linked target genes. List of candidate target genes for the tumor-gained mutated ERα sites based on H3K27ac HiChIP analysis of Ishikawa cells [91]. The structure of the table is the following: *chr_enh/start_enh/end_enh*: genomic position of the ERα binding sites; *count_enh*: enhancer identifier; *target_gene/ensemble_ID*: candidate target gene of enhancer from endometrial cancer H3K27ac

HiChIP data; *enhancer_in_HiChIP_target_gene_promoter_anchor?*: *yes*, enhancer is located within the HiChIP anchor that encompasses the gene TSS—*no*, enhancer is located in the HiChIP anchor distal to the gene; *enhancer_within_3kb_TSS_of_target_gene?*: *yes*, enhancer is located within ±3 kb of the TSS of the gene—*no*, enhancer is not located within 3 kb of the TSS of the gene

Additional file 6: Uncropped images of blot in Fig. 3i

Additional file 7: Uncropped images of blot in Fig. 5e

Additional file 8: Uncropped images of blot in Fig. 5 g

Acknowledgements
We thank members of the Zwart and Bergman labs for valuable feedback, suggestions and input. We would like to acknowledge the Research High Performance Computing (RHPC) facility of the Netherlands Cancer Institute (NKI) to have enabled us to perform all the computations required to analyze the data generated, the NKI Genomics Core Facility for next-generation sequencing and bioinformatics support, the NKI-AVL Core Facility Molecular Pathology and Biobanking for patient samples handling and, the NKI Library for the support in the patient raw data deposition. This publication and the underlying study have been made possible partly based on data that Hartwig Medical Foundation has made available to the study through the Hartwig Medical Database. Further, we would like to thank Hartwig Medical Foundation for facilitating and supporting whole-genome sequencing data generation, data analyses and data interpretation.

Peer review information
Guohong Li and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
Conceptualization: S.G., A.K and W.Z. W.Z. was responsible for project funding. S.G. and A.K. performed Hi-C and 4 C-seq experiments; S.G., K.S. and M.D. performed ChIP/ChIP-seq and immunoblotting experiments; S.S. performed and analyzed the protein DNA-oligo pulldowns experiments; S.G., M.H, T.M.S. and A.A.S. analyzed the omics data; T.A.O. and D.M.G. provided support in the HiChIP and rSNP data analyses; M.V. provided support for the proteomics experiments; L.F.A.W. provided bioinformatics support; F.E.v.L. designed and performed study for patient sample collection. S.G., A.K. and W.Z. wrote the manuscript, with input from all co-authors. All authors read and approved the final manuscript.

Authors' X/Bluesky handles
Bluesky handles: @sgregoricchio.bsky.social/Twitter handles: @s_gregoricchio (Sebastian Gregoricchio). Bluesky handles: @zwartlab.bsky.social (Wilbert Zwart).

Funding
This work was supported by Oncode Institute and ***Saxum Volutum*** . The Zwart, Wessels, and Vermeulen labs are part of the Oncode Institute, which is partly funded by the Dutch Cancer Society. T.O. is supported by a National Health and Medical Research Council of Australia Emerging Leader Investigator Fellowship (GNT1173170).

Data availability
Endometrial healthy and tumor tissues ERα and H3K27ac ChIP-seq, RNA-seq, Hi-C and, 4C-seq raw sequencing data (GRCh37/Hg19 genome build) have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001007240 [91], while endometrial tissue processed data and Ishikawa ChIP-seq data are deposited at the Gene Expression Omnibus (GEO) database under accession number GSE235241 [92]. Raw sequencing data of the RNA-seq performed in Ishikawa cells can be found at GEO database under accession number GSE109892 [31, 93]. ERα ChIP-seq data in breast cancer tissues from Singh et al**.** [33] are accessible in GEO under accession number GSE114737 [94]. Ishikawa ChIP-seq data in the form of optimal Irreproducibility Discovery Rate (IDR) thresholded peaks (GRCh37/Hg19 genome build) were obtained from the ENCODE data portal [50] (Additional file 3: Table S2). The H3K27ac HiChIP data in Ishikawa cells have been obtained from O'Mara et al. [93] accessible in GEO under accession number GSE137936 [96]. Mass spectrometry data have been deposited at the ProteomeXchange Consortium through the PRIDE [97] partner repository with the identifier PXD029822 [98]. The breast and endometrial cancer ATAC-seq, gene expression, progression-free survival and Whole Genome Sequencing (data access authorization for project ID: 36269) used in this this work are in whole or part based upon data generated by the TCGA Research Network [26, 32] (https://www.cancer.gov/tcga).

Declarations

Ethics approval and consent to participate
This study was approved by the local medical ethics committee of the Netherlands Cancer Institute (Institutional Review Board (IRB) reference number: IRBdm19-277) and complies with the ethical principles of the Declaration of Helsinki. All patients provided informed consent for translational studies.

Competing interests
The authors declare no competing interests.

Author details
[1]Division of Oncogenomics, Oncode Institute, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. [2]Division of Molecular Carcinogenesis, Oncode Institute, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. [3]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, Geert Grooteplein Zuid 28,

6525GA Nijmegen, The Netherlands. [4]Cancer Research Program, QIMR Berghofer Medical Research Institute, Locked Bag 2000, Brisbane, QLD 4029, Australia. [5]Division of Molecular Genetics, Oncode Institute, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. [6]Department of Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. [7]Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Received: 10 June 2024 Accepted: 28 April 2025
Published online: 09 May 2025

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Gu B, Shang X, Yan M, Li X, Wang W, Wang Q, et al. Variations in incidence and mortality rates of endometrial cancer at the global, regional, and national levels, 1990–2019. Gynecol Oncol. 2021;161:573–80.
3. Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer Lancet Lond Engl. 2016;387:1094–108.
4. Carlson MJ, Thiel KW, Leslie KK. Past, present, and future of hormonal therapy in recurrent endometrial cancer. Int J Womens Health. 2014;6:429–35.
5. Rodriguez AC, Blanchard Z, Maurer KA, Gertz J. Estrogen signaling in endometrial cancer: a key oncogenic pathway with several open questions. Horm Cancer. 2019;10:51–63.
6. Gunderson CC, Fader AN, Carson KA, Bristow RE. Oncologic and reproductive outcomes with progestin therapy in women with endometrial hyperplasia and grade 1 adenocarcinoma: a systematic review. Gynecol Oncol. 2012;125:477–82.
7. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N. Endometrial cancer Lancet Lond Engl. 2022;399:1412–28.
8. Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, et al. Estrogen receptors: how do they signal and what are their targets. Physiol Rev. 2007;87:905–31.
9. Flach KD, Zwart W. The first decade of estrogen receptor cistromics in breast cancer. J Endocrinol. 2016;229:R43-56.
10. Droog M, Nevedomskaya E, Kim Y, Severson T, Flach KD, Opdam M, et al. Comparative cistromics reveals genomic cross-talk between FOXA1 and ERα in tamoxifen-associated endometrial carcinomas. Cancer Res. 2016;76:3773–84.
11. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol Cell. 2013;52:25–36.
12. Gertz J, Reddy TE, Varley KE, Garabedian MJ, Myers RM. Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. Genome Res. 2012;22:2153–62.
13. van Leeuwen FE, van den Belt-Dusebout AW, van Leeuwen FE, Benraadt J, Diepenhorst FW, van Tinteren H, et al. Risk of endometrial cancer after tamoxifen treatment of breast cancer. The Lancet. 1994;343:448–52.
14. Droog M, Nevedomskaya E, Dackus GM, Fles R, Kim Y, Hollema H, et al. Estrogen receptor α wields treatment-specific enhancers between morphologically similar endometrial tumors. Proc Natl Acad Sci. 2017;114:E1316-25.
15. Chi D, Singhal H, Li L, Xiao T, Liu W, Pun M, et al. Estrogen receptor signaling is reprogrammed during breast tumorigenesis. Proc Natl Acad Sci U S A. 2019;116:11437–43.
16. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat Genet. 2015;47:1346–51.
17. Mazrooei P, Kron KJ, Zhu Y, Zhou S, Grillo G, Mehdi T, et al. Cistrome partitioning reveals convergence of somatic mutations and risk variants on master transcription regulators in primary prostate tumors. Cancer Cell. 2019;36:674-689.e6.
18. Pomerantz MM, Qiu X, Zhu Y, Takeda DY, Pan W, Baca SC, et al. Prostate cancer reactivates developmental epigenomic programs during metastatic progression. Nat Genet. 2020;52:790–9.
19. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012;481:389–93.
20. Layer RM, Pedersen BS, Disera T, Marth GT, Gertz J, Quinlan AR. GIGGLE: a search engine for large-scale integrated genome analysis. Nat Methods. 2018;15:123–6.
21. Robson MI, Ringel AR, Mundlos S. Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D. Mol Cell. 2019;74:1110–22.
22. Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biol. 2017;18:125.
23. de Wit E, Bouwman BAM, Zhu Y, Klous P, Splinter E, Verstegen MJAM, et al. The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature. 2013;501:227–31.
24. Gregoricchio S, Zwart W. snHiC: a complete and simplified snakemake pipeline for grouped Hi-C data analysis. Bioinforma Adv. 2023;3:vbad080.
25. Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature. 2020;578:102–11.
26. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45:1113–20.
27. O'Mara TA, Glubb DM, Amant F, Annibali D, Ashton K, Attia J, et al. Identification of nine new susceptibility loci for endometrial cancer. Nat Commun. 2018;9:3166.
28. Boot A, Ng AWT, Chong FT, Ho S-C, Yu W, Tan DSW, et al. Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. Genome Res. 2020;30:803–13.

29. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497:67–73.
30. O'Hara AJ, Bell DW. The genomics and genetics of endometrial cancer. Adv Genomics Genet. 2012;2012:33–47.
31. Vahrenkamp JM, Yang C-H, Rodriguez AC, Almomen A, Berrett KC, Trujillo AN, et al. Clinical and genomic crosstalk between glucocorticoid receptor and estrogen receptor α in endometrial cancer. Cell Rep. 2018;22:2995–3005.
32. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science. 2018;362:eaav1898.
33. Singh AA, Schuurman K, Nevedomskaya E, Stelloo S, Linder S, Droog M, et al. Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. Life Sci Alliance. 2019 [cited 2023 Apr 22];2. Available from: https://www.life-science-alliance.org/content/2/1/e201800115
34. Bartha Á, Győrffy B. TNMplo.com: a web tool for the comparison of gene expression in normal, tumor and metastatic tissues. Int J Mol Sci. 2021;22:2622.
35. Fudenberg G, Kelley DR, Pollard KS. Predicting 3D genome folding from DNA sequence with Akita. Nat Methods. 2020;17:1111–7.
36. Zwart W, Rondaij M, Jalink K, Sharp ZD, Mancini MA, Neefjes J, et al. Resistance to antiestrogen arzoxifene is mediated by overexpression of cyclin D1. Mol Endocrinol Baltim Md. 2009;23:1335–45.
37. Perou CM, Sørlie T, Eisen MB, Van De Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.
38. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, et al. Genome-wide analysis of estrogen receptor binding sites. Nat Genet. 2006;38:1289–97.
39. Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, et al. Whole-genome cartography of estrogen receptor α binding sites. PLOS Genet. 2007;3: e87.
40. Zhang Y, Bao W, Wang K, Lu W, Wang H, Tong H, et al. SOX17 is a tumor suppressor in endometrial cancer. Oncotarget. 2016;7:76036–46.
41. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a critical determinant of estrogen receptor function and endocrine response. Nat Genet. 2011;43:27–33.
42. Yamaguchi K, Chen X, Oji A, Hiratani I, Defossez P-A. Large-scale chromatin rearrangements in cancer Cancers. 2022;14:2384.
43. Johnstone SE, Reyes A, Qi Y, Adriaens C, Hegazi E, Pelka K, et al. Large-scale topological changes restrain malignant progression in colorectal cancer. Cell. 2020;182:1474-1489.e23.
44. Yang Q, Jiang N, Zou H, Fan X, Liu T, Huang X, et al. Alterations in 3D chromatin organization contribute to tumorigenesis of EGFR-amplified glioblastoma. Comput Struct Biotechnol J. 2022;20:1967–78.
45. Sabarinathan R, Mularoni L, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature. 2016;532:264–7.
46. Purcell DJ, Jeong KW, Bittencourt D, Gerke DS, Stallcup MR. A distinct mechanism for coactivator versus corepressor function by histone methyltransferase G9a in transcriptional regulation. J Biol Chem. 2011;286:41963–71.
47. Zhang X, Peng D, Xi Y, Yuan C, Sagum CA, Klein BJ, et al. G9a-mediated methylation of ERα links the PHF20/MOF histone acetyltransferase complex to hormonal gene expression. Nat Commun. 2016;7:10810.
48. Chaturvedi C-P, Hosey AM, Palii C, Perez-Iratxeta C, Nakatani Y, Ranish JA, et al. Dual role for the methyltransferase G9a in the maintenance of beta-globin gene transcription in adult erythroid cells. Proc Natl Acad Sci U S A. 2009;106:18303–8.
49. Chaturvedi C-P, Somasundaram B, Singh K, Carpenedo RL, Stanford WL, Dilworth FJ, et al. Maintenance of gene silencing by the coordinate action of the H3K9 methyltransferase G9a/KMT1C and the H3K4 demethylase Jarid1a/KDM5A. Proc Natl Acad Sci U S A. 2012;109:18845–50.
50. Korch C, Spillman MA, Jackson TA, Jacobsen BM, Murphy SK, Lessey BA, et al. DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. Gynecol Oncol. 2012;127:241–8.
51. Gregoricchio S, Polit L, Esposito M, Berthelet J, Delestré L, Evanno E, et al. HDAC1 and PRC2 mediate combinatorial control in SPI1/PU.1-dependent gene repression in murine erythroleukaemia. Nucleic Acids Res. 2022;50:7938–58.
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009;25:1754–60.
53. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.
54. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
55. Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, et al. Uniform, optimal signal processing of mapped deep-sequencing data. Nat Biotechnol. 2013;31:615–22.
56. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: identification of problematic regions of the genome. Sci Rep. 2019;9:9354.
57. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf Engl. 2010;26:139–40.
58. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160-165.
59. Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics. 2021;37:422–3.
60. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun. 2018;9:189.
61. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31:2382–3.
62. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics. 2010;11:165.
63. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43:W39-49.

64. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022;50:D165-73.
65. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 2004;32:D91-94.
66. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50:W216-21.
67. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.
68. Ambrosini G, Groux R, Bucher P. PWMScan: a fast tool for scanning entire genomes with a position-specific weight matrix. Bioinformatics. 2018;34:2483–4.
69. Donaldson-Collier MC, Sungalee S, Zufferey M, Tavernari D, Katanayeva N, Battistello E, et al. EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. Nat Genet. 2019;51:517–28.
70. Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 2020;48:W177-84.
71. Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R, et al. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 2018;46:W11-6.
72. Chakraborty A, Wang JG, Ay F. dcHiC detects differential compartments across multiple Hi-C datasets. Nat Commun. 2022;13:6827.
73. van der Weide RH, van den Brand T, Haarhuis JHI, Teunissen H, Rowland BD, de Wit E. Hi-C analyses with GENOVA: a case study with cohesin variants. NAR Genomics Bioinforma. 2021;3:lqab040.
74. Wang S, Lee S, Chu C, Jain D, Kerpedjiev P, Nelson GM, et al. HiNT: a computational method for detecting copy number variations and translocations from Hi-C data. Genome Biol. 2020;21:73.
75. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics. 2017;18:287.
76. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575:210–6.
77. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. JCO Precis Oncol. 2017;2017:PO.17.00011.
78. Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. Curr Protoc Bioinforma. 2014;47:11.12.1-34.
79. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
80. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47:D941-7.
81. Ahmed M, Sallari RC, Guo H, Moore JH, He HH, Lupien M. Variant Set Enrichment: an R package to identify disease-associated functional genomic regions. BioData Min. 2017;10:9.
82. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, Abecasis GR, Steering committee, Altshuler DM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
83. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10–2.
84. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
85. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
86. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. The Innovation. 2021;2:100141.
87. Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W. 4C-seq from beginning to end: a detailed protocol for sample preparation and data analysis. Methods San Diego Calif. 2020;170:17–32.
88. Vermeulen M. Identifying chromatin readers using a SILAC-based histone peptide pull-down approach. Methods Enzymol. 2012;512:137–60.
89. Makowski MM, Willems E, Fang J, Choi J, Zhang T, Jansen PWTC, et al. An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations. Proteomics. 2016;16:417–26.
90. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367–72.
91. S. Gregoricchio, A. Kojic, M. Hoogstraat, K. Schuurman, S. Stelloo, T.M. Severson, T.A. O'Mara, M. Droog, A.A. Singh, D.M. Glubb, L.F.A. Wessels, M. Vermeulen, F.E. van Leeuwen, W. Zwart. Epigenetic plasticity in endometrial tumorigenesis demarcates non-coding somatic mutations and 3D-genome alterations boosting tumor progression. European Genome-phenome Archive; 2025. Available from: https://ega-archive.org/studies/EGAS00001007240
92. S. Gregoricchio, A. Kojic, M. Hoogstraat, K. Schuurman, S. Stelloo, T.M. Severson, T.A. O'Mara, M. Droog, A.A. Singh, D.M. Glubb, L.F.A. Wessels, M. Vermeulen, F.E. van Leeuwen, W. Zwart. Epigenetic plasticity in endometrial tumorigenesis demarcates non-coding somatic mutations and 3D-genome alterations boosting tumor progression. Gene Expression Omnibus; 2025. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE235241
93. J. Gertz, J. Vahrenkamp. Clinical and genomic crosstalk between glucocorticoid receptor and estrogen receptor α in endometrial cancer [RNA-seq]. Gene Expression Omnibus; 2018. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109892
94. A.A. Singh, K. Schuurman, E. Nevedomskaya, S. Stelloo, S. Linder, M. Droog, Y. Kim, J. Sanders, H. van der Poel, A.M. Bergman, L.F.A. Wessels, W. Zwart. Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. Gene Expression Omnibus; 2018. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114737

95. O'Mara TA, Spurdle AB, Glubb DM, Endometrial Cancer Association Consortium. Analysis of promoter-associated chromatin interactions reveals biologically relevant candidate target genes at endometrial cancer risk loci. Cancers. 2019;11:1440.
96. DM Glubb, TA O'Mara. Analysis of promoter-associated chromatin interactions reveals biologically relevant candidate target genes at endometrial cancer risk loci. Gene Expression Omnibus; 2020. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137936
97. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50:D543-52.
98. S. Gregoricchio, A. Kojic, M. Hoogstraat, K. Schuurman, S. Stelloo, T.M. Severson, T.A. O'Mara, M. Droog, A.A. Singh, D.M. Glubb, L.F.A. Wessels, M. Vermeulen, F.E. van Leeuwen, W. Zwart. Functional analysis of somatic variants found in endometrial tumor-specific ER-alpha enhancers. ProteomeXchange Consortium; 2025. Available from: https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD029822

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.