Enhancer plasticity in endometrial tumorigenesis demarcates non-coding driver mutations and alterations in 3D genome organization to stimulate oncogene expression

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 van Leeuwen[6], Wilbert Zwart[1]

[1] Division of Oncogenomics, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[2] Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, Nijmegen, The Netherlands
[4] Cancer Research Program, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[5] Division of Molecular Genetics, Oncode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands
[6] Department of Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[#] Authors equally contributed to this work

ABSTRACT

The incidence and mortality of Endometrial Cancer (EC) is on the rise. 85% of ECs depend on Estrogen Receptor alpha (ERα) for proliferation, but little is known about its transcriptional regulation in these tumors.

We generated epigenomics and Hi-C data streams 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 WGS data from metastatic samples revealed a striking enrichment of non-coding somatic mutations at tumor-enriched ERα sites. Through machine learning-based predictions and interaction proteomics analyses, we identified an enhancer mutation which alters 3D genome organization, impairing recruitment of the transcriptional repressor EHMT2/G9a/KMT1C, thereby alleviating transcriptional repression of *ESR1* in EC.

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

INTRODUCTION

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 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 ovarian function in young patients. Furthermore, the estrogen competitive ERα antagonist tamoxifen is also prescribed to patients with endometrial cancer, often alternated with PR agonists (reviewed in ref [6]). 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[7,8]. To date, most literature on promoter-enhancer communication and ERα activity has been focused on breast cancer[8], and is far less understood in endometrial tumors. The observation that only around 30% of ERα sites are shared between breast cancer and endometrial cancer[9–11] highlights highly context dependent chromatin binding of ERα. Tamoxifen – prescribed in the treatment of breast cancer – is a risk factor for endometrial tumorigenesis[12]. 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[13]. These observations highlighted that the endometrial cancer epigenome is not fixed, but rather dynamically affected by oncogenic factors.

Both ERα in breast tissue[14], and Androgen Receptor (AR) in prostate tissue[15–17], 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[16]. In endometrial cancer, such studies thus far not reported 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 leads to EC progression. We identified substantial reprogramming of ERα on the genome upon tumorigenesis, which coincided with the occurrence of somatic variants in metastatic samples. In particular, we discovered an EC-specific *ESR1* enhancer which is target of somatic variants in metastatic EC, loosing capacity to recruit lysine methyl-transferase G9a/EHMT2/KMT1C when mutated, to enhance ERα expression in metastatic disease. Cumulatively, we show that non-coding mutations in endometrial cancer can 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 previous systemic therapy, as well as four fresh-frozen samples from women with pathologically normal endometrial tissue (**Fig. 1a** and **Extended Data Fig. 1a**, for clinicopathological characteristics see **Extended Data Table 1**). 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 1,399-51,888) and 44,824 H3K27ac peaks (range 8,791-71,898) in our samples. The ChIP-seq library size was comparable among the different samples for both targets (**Extended Data Fig. 1b**). However, 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 asses data quality and reproducibility, we performed Principal Component Analysis (PCA) on the peak intensities. These analyses revealed a clustering of the samples by ChIP target (first component) and tissue type (second component) (**Extended Data Fig. 1c**).

Differential binding analysis (*DiffBind*[18]) for ERα between tumor and normal samples revealed over 10,292 differentially-bound genomic locations (FDR ≤ 0.05 & $|\log_2(\text{FoldChange})| \geq 1$) (**Fig. 1d-e** and **Extended Data Fig. 1d**), of which 6,488 ERα sites were lost and 3,804 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 confirmed the selective enrichment of signals between samples, with high reproducibility (**Fig. 1f**). As ERα acts as enhancer-selective transcription factor, 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**).

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[9,13]) and others (cell lines[10]), tumor-enriched ERα sites were mostly found at distal intergenic and intronic regions (**Fig. 1h**). 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 (**Extended Data Fig. 1e**). 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 1i**).

Differential motif enrichment suggests selective enrichment of ERα-interaction partners for different regions. To test this hypothesis, we integrated publicly available ChIP-seq data from the Ishikawa endometrial cancer cell line (**Fig. 1j** and **Extended Data Table 1**), revealing striking differences between transcription factors on chromatin selectivity 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 $P = 3.15\times10^{-21}$), while ERα peaks from estradiol-stimulated cells showed significant overlap with the tumor-enriched ERα peaks (adjusted Fisher's exact $P = 4.60\times10^{-59}$, **Fig. 1j**). In addition, both FOXM1 and FOXA1 binding sites in Ishikawa showed significant overlaps with our tumor enriched regions (adjusted Fisher's exact $P = 3.37\times10^{-21}$ and $P = 6.25\times10^{-57}$, respectively, **Fig. 1k**).

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[19]. 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 **Extended Data Table 1**). Hi-C is particularly suitable to study translocations in cancer[20], which were not

observed in any of the tumor samples we analyzed (**Extended Data Fig. 2a**). We did however observe a distinct clustering of the healthy tissues from the tumor specimens, based on 3D genome organization (**Extended Data Fig. 2b**). In addition, tumor samples displayed a higher Hi-C contact heterogeneity in comparison to the more correlated healthy tissues (**Extended Data Fig. 2b**). Next, we quantified the Relative Hi-C Contact Probability (RCP) as function of the contact distances. Interestingly, tumors displayed an increased probability of short-range interactions (<2Mb) concomitant with a loss of long-range chromatin contacts (**Fig. 2b-d** and **Extended Data Fig. 2c-d**). Furthermore, analyzing the distribution of the loop distances (**Fig. 2e**), 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 next wondered whether this loop shortening in tumor samples was associated with changes in chromatin compartmentalization. Analysis of compartment polarization (**Fig. 2b,f-g** and **Extended Data Fig. 2e-f**) 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) where slightly stronger in tumor tissues (**Fig. 2g** and **Extended Data Fig. 2f**). Overall, around one third of compartments (A: 565/1431 (39.5%); B: 440/1306 (33.6%)) underwent class switching in tumor samples (**Fig. 2h**). To investigate whether this compartment class switch was occurring specifically in compartments where ERα was bound in a tissue type-specific manner, we overlapped the compartments with the differential ERα occupied regions upon tumorigenesis (**Fig. 2i**). 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: $P = 1.317\times10^{-56}$, OR 0.20; A-to-B depleted-vs-enriched: $P = 6.013\times10^{-11}$, OR 1.87) (**Fig. 2j** and **Extended Data Fig. 2g**).

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 short-range chromatin interactions upon tumor development, at the expense of long-range contacts.

Somatic mutations are enriched at tumor-gained ERα sites

Somatic mutations in cancer are often found enriched at non-coding regulatory elements[21]. To determine whether that is also the case in endometrial cancer, we analyzed Whole Genome Sequencing (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. 3a** and **Extended Data Table 3**). 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. 3b**). To assess WGS data quality and explore additional genomic features of our samples, we investigated mutational signatures, counts, and frequently-mutated genes (**Fig. 3c-d**). 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[22]. Cumulatively, these results confirm previously-described enrichment of tumor-intrinsic[23,24] and treatment-induced[22] 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[23].

As somatic mutations in metastatic samples were enriched in non-coding regions, we integrated these somatic mutation data with tissues-type enriched ERα ChIP-seq data (**Fig. 3e**). Interestingly, differentially bound ERα regions were significantly enriched for somatic mutations (11.9%, 1,221/10,292), relative to non-differential ERα-bound sites (8%, 5,881/73,312; Fisher's exact $P = 5.64\times10^{-36}$, 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 (14.6% vs 10.1%, Fisher's exact $P = 2.98\times10^{-11}$, OR 1.51, 95% CI 1.34-1.71, **Fig. 3e**). In most instances (n=1,002), 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 ≤ 0.01) bearing mutations with H3K27ac HiChIP chromatin looping data in Ishikawa cells (**Extended Data Table 4**), aimed to couple putative enhancrs 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 8h of estradiol-stimulation[25] (**Fig. 3f**). Interestingly, *ESR1* (encoding ERα) was represented as well, being downregulated at both RNA and protein level (**Fig. 3f-g** and **Extended Data Fig. 3a**) and harboring 3 metastatic somatic mutations in two *cis* regulatory elements upstream of the *ESR1* locus, hereafter referred as '*Enhancer 1*' and '*Enhancer 2*' (**Fig. 4a** and **Extended Data Fig. 3b**). Of note, high *ESR1* expression is associated with favorable outcome in primary endometrial cancer patients (TCGA[26]) (**Fig. 3h**)**.** 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[27]) and ERα ChIP-seq signal (our previous work[28]) (**Extended Data Fig. 3b-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* – targets of somatic mutations – 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.

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-point (**Fig. 4c**), showing tumor-specific gained interactions of newly ERα-engaged enhancers with the *ESR1* promoter (**Extended Data Fig. 2h**).

To identify which *ESR1* enhancer mutation had an higher probability to impact *ESR1* gene regulation and potentially drive tumor progression, we employed the *Akita*[29] 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) 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 the center of the ERα-bound site at the *Enhancer 1* – containing either the reference or mutant sequence – with a ±25bp 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**). We then compared these values with the distribution of the correlation of *ESR1* expression against the full transcriptome, after which the genes most-strongly correlating positively (*POLK*, *ZBTB21*, *XPA*, *SIN3A*) or negatively (*GTF2IRD1*, *EHMT2*, *ZNF768*) with *ESR1* expression, were selected (**Extended Data Fig. 4a**). For these genes, expression was analyzed in relation to the progression-free survival probability of endometrial cancer patients (TCGA[26]) (**Fig. 5d** and **Extended Data Fig. 4b**). 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 **Extended Data Fig. 4b**).

EHMT2/G9a interacts with ERα (**Fig. 5e** and **Extended Data Fig. 4c**), based on immunoprecipitation experiments. In line with this, EHMT2/G9a was recruited to the *ESR1 Enhancer 1* and promoter upon estradiol treatment, as assessed by ChIP-qPCR in Ishikawa cells (**Fig. 5f**). Furthermore, siRNA-mediated knock-down of EHMT2/G9a in Ishikawa cell lines, resulted in an increase of ERα protein expression (**Fig. 5g** and **Extended Data Fig. 4d**), confirming the reciprocal relationship between these proteins.

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 a ERα-dependent manner.

DISCUSSION

Despite the central role of ERα in the endometrial cancer development and progression[5], the molecular actions of ERα in this cancer type remain largely unknown.

Though 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. 5h**): i) in healthy endometrial tissue, ERα regulates genes involved in cell cycle and DNA damage regulation; ii) during tumorigenesis, ERα re-localizes to the tumor-specific *ESR1* enhancer where recruits and interacts with EHMT2/G9a which acts as a co-repressor to reduce *ESR1* expression; iii) during tumor progression, the *ESR1* enhancer acquires a somatic mutation that may impair EHMT2/G9a chromatin occupancy to allow for enhanced *ESR1* expression.

Notably, our analysis of the ERα cistrome in both normal and neoplastic endometrial tissues revealed limited overlap of chromatin occupancy between the two tissue states. 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. Indeed, ERα has been reported to primarily occupy promoter-distal *cis*-regulatory elements rather than promoters[9,13,30,31]. However, the molecular biology of this nuclear receptor has been studied mainly using cancer models, introducing therefore a potential cancer-related bias. Indeed, our results highlight an extended re-programming of the ERα regulation in tumors involving a promoter-to-enhancer shift in ERα chromatin occupancy.

Our data, combined with previous findings showing that cell-specific ERα sites lack high-affinity estrogen responsive elements (EREs)[10], 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 (lacking 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*[23], which acts as tumor suppressor[32]. 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. Indeed, FOXA1 is known to facilitate ERα binding through its pioneer activity in breast cancer[33], and we previously proposed FOXA1 to serve a comparable role in endometrial cancer[9]. 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. 1j-k**).

Enhancers interact with promoter regions to regulate gene expression, facilitated through spatial rearrangement of the chromatin structure[19]. Performing Hi-C analyses in healthy and tumor tissues, we observed a global enrichment of short-range (<2Mb) chromatin interactions, at the expense of long-range (>2Mb) contacts, and compartment depolarization in tumor samples. Chromatin compartment destabilization is observed in many tumor types[34] (e.g., colorectal cancer[35] and glioblastoma[36]). 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 somatic mutations preferentially occur at non-coding regions of the genome[21], 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[16].

While the interaction between ERα and EHMT2/G9a has already been reported in breast[37,38], EHMT2/G9a is considered to be an ERα co-activator in that cellular context. However, studies in eryhtroleukemia[39,40] 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, but further investigation would be required to define what determines the repressive transcriptional effect of EHMT2/G9a in the interplay with ERα in this tumor type.

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. This study may serve as

blueprint for studies in other hormone-driven cancer types, in which such complex connections are yet to be identified.

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 haematoxylin 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.

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 endometrial cancer cell lines (Merck Sigma Aldrich) have been cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% Fetal Bovine Serum (FBS-12A, Capricorn Scientific) and penicillin/streptomycin (100µg/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) supplied with 5% dextran-coated charcoal (DCC) stripped FBS, 2mM L-glutamine (Gibco), and penicillin/streptomycin (100µg/mL, Gibco), then stimulated with 10nM 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 100nM DMSO-solubilized Fulvestrant (MedChemExpress, #HY-13636).

ChIP-qPCR and ChIP-seq library preparation

Chromatin Immunoprecipitation of ERα (SantaCruz #sc-543, 5µg/IP) and H3K27ac (ActiveMotif #39133, 5µg/IP) were performed as previously described[28] employing the combination of DSG and Formaldehyde to perform the crosslinking. The same protocol, but using only Formaldehyde for the crosslinking, have been used to perform EHMT2/G9a (Abcam #ab133482, 5µg/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[41]. Reads with a mapping quality (MAPQ) < 20 were removed from further analysis. Enrichment over input control was determined using both *MACS2*[42] and *DFilter*[43]. Only peaks identified by both methods, and not overlapping with the ENCODE GRCh37/Hg19 blacklisted regions[44], were retained.

Differentially bound regions between tumor and healthy samples were identified using *DiffBind*[18] with *EdgeR*[45] mode at an FDR < 0.05. RPGC-normalized (Reads Per Genomic Content; *bamCoverage*[46]) ChIP-seq signal at peaks was visualized using *deepTools*[46], while genomic tracks and average density profiles were generated with *Rseb*[47] R-package (https://github.com/sebastian-gregoricchio/Rseb) or *pyGenomeTracks*[48,49].

Peak genomic annotation was performed using *Homer*[50], while for motif enrichment in peak regions we used *AME* v5.5.05[51] algorithm from the *MEME* suite[52] using the JASPAR database[53,54]. 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* R package.

Analyses for Gene Ontology biological process (GO-BP) enrichments for tumor-depleted ERα promoter-bound genes were performed using DAVID[55] (v6.8) and employing tumor-enriched ERα promoter-bound genes as background data set.

Hi-C library preparation and data processing

Flash-frozen primary tissues have been processed as described in **Fig. 2a**. Briefly, 30µm slices are cross-linked, washed, electronically homogenized and filtered using a 75µm cell strainer and collected in a 1.5mL microcentrifuge loBind tube. Then, Hi-C single-index library preparation was performed as previously described[56] 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 75bp paired-end setup. De-multiplexed fastq data were analyzed at 2kb, 10kb, 40kb, 100kb and 500kb resolution using the *snHiC*[57] v0.1.1 pipeline (https://github.com/sebastian-gregoricchio/snHiC) using default parameters and the

Hg19/GRCh37 genome assembly. Downstream analyses have been performed using *GENOVA*[58] v1.0.1 (Hi-C contact matrices and Relative Contact Probabilities), *HiCExplorer*[59,60] v3.7.2 (compartment polarization, loops distance, Hi-C correlations, TAD detection, long/short contacts ratio), *intervene*[61] (upset plot and compartment-ChIP peaks overlaps), *dcHiC*[62] (compartments detection and differential compartment analyses), *pyGenomeTracks*[48,49] (data visualization).

Whole Genome Sequencing (WGS) analyses

Sampling, sequencing and initial computational steps have been previously described extensively by Priestley *et al.*[63] and performed at the central sequencing facility at the Hartwig Medical Foundation.

Mutational signatures were determined using the *deconstructSigs*[64] R package, using the COSMIC mutational signatures v3 database[65]. 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[66]. Overlap between somatic variants and ChIP-seq regions was determined using *bedtools* v2.25.0 intersect and *multiinter*[67].

RNA-seq and differential gene expression analyses

Ishikawa gene expression data was obtained from GEO, accession numbers GSE109892[25]. Differential gene expression between estradiol and DMSO treated cells was computed using *DESeq2*[68].

Circularized Chromosome Conformation Capture sequencing (4C-seq)

For each sample, 10-15 30µM-thick slices of flash frozen tissue were cross-linked under rotation at room temperature for 20 minutes in 5mL of 1× PBS supplemented with 2% FBS and 2% Methanol-free Formaldehyde. The cross-linked slices have been washed twice in 10mL of ice-cold 1× PBS before to be resupended in 0.5-1mL of ice-cold 1× PBS in 1.5mL DNA-loBind microcentrifuge tubes. Tissues have been electronically homogenized and lysates filtered using a 75µm cell strainer. Then, experiments of 4C-seq were performed as previously described[69]. 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[69] 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 75bp Single-End reads setup. Fastq files have been demultiplexed by Cutadapt[70] and mapped on Hg19/GRCh37 genome assembly and signal normalized by *pipe4C*[69] 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 have 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*[47] (v0.3.0) (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 immunoprecipiated proteins, elution buffer (1% SDS, 0.1M $NaHCO_3$) and Laemmli buffer, supplied with 100mM DTT (Dithiothreitol, Merck Sigma Aldrich), were added to input samples and beads, and then boiled at 95°C for 30min. For whole-cell extracts, cells have been lysed using 2× Laemmli buffer supplied with 100mM DTT (Dithiothreitol, Merck Sigma Aldrich), boiled at 95°C for 10min and then DNA was sheared by sonication (Bioruptor® Pico, Diagenode; 4 cycles of 30s ON + 90s OFF). In both cases, proteins were then resolved by SDS-PAGE and transferred to a 0.22µm nitrocellulose membrane (BioRad). Membranes were incubated for 2h 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® 800CW (926-32212, LI-COR Biosciences, 1:10,000 in BS), donkey-α-rabbit IRDye® 800CW (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 1h. 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*[29] 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[71]. The ~50bp oligonucleotide probes encompassing the SNP were ordered with the forward strand containing a 5'-biotin moiety (Integrated DNA Technologies) (Fw_WT: 5'-/5Biosg/ACTGTGGAAACTGGAAGCTGTTCTTGGACTATTTCGCAACACTTTTCTCC-3'; Rv_WT: 5'-GGAGAAAAGTGTTGCGAAATAGTCCAAGAACAGCTTCCAGTTTCCACAGT-3'; Fw_SNP: 5'-/5Biosg/ACTGTGGAAACTGGAAGCTGTTTGGACTATTTCGCAACACTTTTCTCCTC-3', Rv_SNP: (5'-GAGGAGAAAAGTGTTGCGAAATAGTCCAAACAGCTTCCAGTTTCCACAGT-3'; mutated site is underlined). DNA affinity purifications, on-bead trypsin digestion and dimethyl labeling were performed as described[72]. Matching light and medium labelled samples were then combined and analyzed using a gradient from 7 to 30% buffer B in Buffer A over 44min, followed by a further increase to 95% in the next 16min at flow rate of 250nL/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[72,73]. 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 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.

DATA AVAILABILITY

Endometrial healthy and tumor tissues ERα and H3K27ac ChIP-seq, Hi-C and, 4C-seq sequencing data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001007240. Raw sequencing data of the RNA-seq performed in Ishikawa cells can be found at the Gene Expression Omnibus (GEO) database under accession number GSE109892[25]. ERα ChIP-seq data in breast cancer tissues from Singh *et al.*[28] are accessible in GEO under accession number GSE114737. 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[74] (**Extended Data Table 1**). The H3K27ac HiChIP data in Ishikawa cells have been obtained from O'Mara *et al.*[75] accessible in GEO under accession number GSE137936. Mass spectrometry data have been deposited at the ProteomeXchange Consortium through the PRIDE[76] partner repository with the identifier PXD029822. The breast and endometrial cancer ATAC-seq, gene expression and progression-free survival used in this this work are in whole or part based upon data generated by the TCGA Research Network[26,27] (https://www.cancer.gov/tcga).

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, and the NKI-AVL Core Facility Molecular Pathology and Biobanking for patient samples handling.

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.

AUTHOR 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 4C-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 data analyses; M.V. provided support for the proteomics experiments; L.F.A.W. provided bioinformatics support; F.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.

FUNDING

This work was supported by Oncode Institute and *Saxum Volutum*. The Vermeulen lab is 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).

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Gu, B. *et al.* Variations in incidence and mortality rates of endometrial cancer at the global, regional, and national levels, 1990-2019. *Gynecol. Oncol.* **161**, 573–580 (2021).
3. Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N. & Darai, E. Endometrial cancer. *Lancet Lond. Engl.* **387**, 1094–1108 (2016).
4. Carlson, M. J., Thiel, K. W. & Leslie, K. K. Past, present, and future of hormonal therapy in recurrent endometrial cancer. *Int. J. Womens Health* **6**, 429–435 (2014).
5. Rodriguez, A. C., Blanchard, Z., Maurer, K. A. & Gertz, J. Estrogen Signaling in Endometrial Cancer: a Key Oncogenic Pathway with Several Open Questions. *Horm. Cancer* **10**, 51–63 (2019).
6. Crosbie, E. J. *et al.* Endometrial cancer. *Lancet Lond. Engl.* **399**, 1412–1428 (2022).
7. Heldring, N. *et al.* Estrogen receptors: how do they signal and what are their targets. *Physiol. Rev.* **87**, 905–931 (2007).
8. Flach, K. D. & Zwart, W. The first decade of estrogen receptor cistromics in breast cancer. *J. Endocrinol.* **229**, R43-56 (2016).
9. Droog, M. *et al.* Comparative Cistromics Reveals Genomic Cross-talk between FOXA1 and ERα in Tamoxifen-Associated Endometrial Carcinomas. *Cancer Res.* **76**, 3773–3784 (2016).
10. Gertz, J. *et al.* Distinct Properties of Cell-Type-Specific and Shared Transcription Factor Binding Sites. *Mol. Cell* **52**, 25–36 (2013).
11. Gertz, J., Reddy, T. E., Varley, K. E., Garabedian, M. J. & Myers, R. M. Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome Res.* **22**, 2153–2162 (2012).
12. Leeuwen, F. E. van *et al.* Risk of endometrial cancer after tamoxifen treatment of breast cancer. *The Lancet* **343**, 448–452 (1994).
13. Droog, M. *et al.* Estrogen receptor α wields treatment-specific enhancers between morphologically similar endometrial tumors. *Proc. Natl. Acad. Sci.* **114**, E1316–E1325 (2017).
14. Chi, D. *et al.* Estrogen receptor signaling is reprogrammed during breast tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11437–11443 (2019).
15. Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
16. Mazrooei, P. *et al.* Cistrome Partitioning Reveals Convergence of Somatic Mutations and Risk Variants on Master Transcription Regulators in Primary Prostate Tumors. *Cancer Cell* **36**, 674-689.e6 (2019).
17. Pomerantz, M. M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat. Genet.* **52**, 790–799 (2020).
18. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
19. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol. Cell* **74**, 1110–1122 (2019).
20. Harewood, L. *et al.* Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* **18**, 125 (2017).
21. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9362–9367 (2009).
22. Boot, A. *et al.* Characterization of colibactin-associated mutational signature in an Asian oral squamous cell carcinoma and in other mucosal tumor types. *Genome Res.* **30**, 803–813 (2020).
23. Cancer Genome Atlas Research Network *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
24. O'Hara, A. J. & Bell, D. W. The genomics and genetics of endometrial cancer. *Adv. Genomics Genet.* **2012**, 33–47 (2012).
25. Vahrenkamp, J. M. *et al.* Clinical and Genomic Crosstalk between Glucocorticoid Receptor and Estrogen Receptor α In Endometrial Cancer. *Cell Rep.* **22**, 2995–3005 (2018).

26. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).

27. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).

28. Singh, A. A. *et al.* Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. *Life Sci. Alliance* **2**, (2019).

29. Fudenberg, G., Kelley, D. R. & Pollard, K. S. Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods* **17**, 1111–1117 (2020).

30. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289–1297 (2006).

31. Lin, C.-Y. *et al.* Whole-Genome Cartography of Estrogen Receptor α Binding Sites. *PLOS Genet.* **3**, e87 (2007).

32. Zhang, Y. *et al.* SOX17 is a tumor suppressor in endometrial cancer. *Oncotarget* **7**, 76036–76046 (2016).

33. Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a critical determinant of Estrogen Receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).

34. Yamaguchi, K., Chen, X., Oji, A., Hiratani, I. & Defossez, P.-A. Large-Scale Chromatin Rearrangements in Cancer. *Cancers* **14**, 2384 (2022).

35. Johnstone, S. E. *et al.* Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* **182**, 1474-1489.e23 (2020).

36. Yang, Q. *et al.* Alterations in 3D chromatin organization contribute to tumorigenesis of EGFR-amplified glioblastoma. *Comput. Struct. Biotechnol. J.* **20**, 1967–1978 (2022).

37. Purcell, D. J., Jeong, K. W., Bittencourt, D., Gerke, D. S. & Stallcup, M. R. A distinct mechanism for coactivator versus corepressor function by histone methyltransferase G9a in transcriptional regulation. *J. Biol. Chem.* **286**, 41963–41971 (2011).

38. Zhang, X. *et al.* G9a-mediated methylation of ERα links the PHF20/MOF histone acetyltransferase complex to hormonal gene expression. *Nat. Commun.* **7**, 10810 (2016).

39. Chaturvedi, C.-P. *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.* **106**, 18303–18308 (2009).

40. Chaturvedi, C.-P. *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.* **109**, 18845–18850 (2012).

41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).

42. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

43. Kumar, V. *et al.* Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31**, 615–622 (2013).

44. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

45. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).

46. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).

47. Gregoricchio, S. *et al.* HDAC1 and PRC2 mediate combinatorial control in SPI1/PU.1-dependent gene repression in murine erythroleukaemia. *Nucleic Acids Res.* **50**, 7938–7958 (2022).

48. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).

49. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).

50. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

51. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).

52. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39-49 (2015).

53. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).

54. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91-94 (2004).

55. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).

56. Donaldson-Collier, M. C. *et al.* EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. *Nat. Genet.* **51**, 517–528 (2019).

57. Gregoricchio, S. & Zwart, W. snHiC: a complete and simplified snakemake pipeline for grouped Hi-C data analysis. *Bioinforma. Adv.* **3**, vbad080 (2023).

58. van der Weide, R. H. *et al.* Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genomics Bioinforma.* **3**, lqab040 (2021).

59. Wolff, J. *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.* **48**, W177–W184 (2020).

60. Wolff, J. *et al.* Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).

61. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).

62. Chakraborty, A., Wang, J. G. & Ay, F. dcHiC detects differential compartments across multiple Hi-C datasets. *Nat. Commun.* **13**, 6827 (2022).

63. Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216 (2019).

64. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

65. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).

66. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**, PO.17.00011 (2017).

67. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12.1-34 (2014).

68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

69. Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods San Diego Calif* **170**, 17–32 (2020).

70. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

71. Vermeulen, M. Identifying chromatin readers using a SILAC-based histone peptide pull-down approach. *Methods Enzymol.* **512**, 137–160 (2012).

72. Makowski, M. M. *et al.* An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations. *Proteomics* **16**, 417–426 (2016).

73. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

74. Korch, C. *et al.* DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. *Gynecol. Oncol.* **127**, 241–248 (2012).

75. O'Mara, T. A., Spurdle, A. B., Glubb, D. M., & Endometrial Cancer Association Consortium. Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci. *Cancers* **11**, 1440 (2019).

76. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).

760

MAIN FIGURE LEGENDS

Fig. 1 | ERα cistrome changes upon tumorigenesis.

A. Schematic workflow of the multi-omics approach applied in this work.

B/C. 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.

E. MA-plot of differential binding analyses performed on the consensus ERα ChIP-seq peaks (Healthy vs Tumor). Differential binding sites (FDR ≤ 0.05 & $|\log_2(\text{FoldChange})| \geq 1$) are highlighted in pink.

F/G. 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. Bar plot with counts per genomic localization of tumor-depleted (blue) and tumor-enriched (orange) ERα consensus peaks.

I. 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-value).

J. 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.

K. Heatmap of the percentage of tumor-depleted or tumor-enriched ERα consensus peaks overlapping with each ChIP-seq target in Ishikawa cells from **j**. Ranking is performed by descending number of overlaps in tumor-depleted peaks.

Fig. 2 | 3D genome landscape is remodeled during tumorigenesis.

A. Schematic of the Hi-C library preparation starting from fresh frozen tissues. Ten 30µm-thick slices of flash-frozen tissue are cross-linked 25min in 2% Formaldehyde. Then, electronically homogenized tissues are filtered using a 75µm cell strainer and subjected to Hi-C library preparation. Hi-C libraries are then sequenced and resulting reads are analyzed by the *snHiC* pipeline.

B. On the left, 40kb-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 (40kb resolution).

D. Violin plot of the distribution of the short-range (<2Mb) over long-rage (>2Mb) Hi-C contacts ratio in each individual sample (40kb resolution). For each comparison the Wilcoxon's test *P*-value is indicated.

E. Distribution of the Hi-C loop distance (10kb resolution) for loops overlapping with tumor-depleted, tumor-enriched or shared ERα consensus peaks in combined healthy (blue) or tumor (orange) endometrial tissues.

F. Compartment polarization ratio (100kb resolution), defined as (AA + BB) / (AB + BA), for each individual sample.

G. Top: saddle plot of A/B compartments interactions as computed in **F** for each individual sample. Bottom: difference of saddle-score with the reference H.005.A2 (healthy tissue), where orange indicates an higher score in tumor samples while purple an higher score in healthy tissues.

H. Sankey plot depicting the proportion of compartment state transition (100kb resolution) of healthy compartments (left) upon tumorigenesis (right).

I. Upset plot showing the overlaps between different compartment transition states (100kb resolution) in healthy compared to tumor as in **H** (A-to-A, B-to-B, A-to-B, B-to-A) and different categories of ERα consensus peaks.

J. Stacked bar plot of the proportion of compartment transition status as in **H** for individual genomic bins overlapping with only one of the different categories of ERα consensus peaks.

Fig. 3 | Tumor-enriched ERα binding sites represent non-coding regions target of somatic mutation in metastatic endometrial cancer tumor.

A. Scheme of the number metastases and metastatic site, for all metastatic samples used for the WGS analyses.

B/C.Genomic localization (**b**) and mutational signature (**c**) distribution of the somatic mutations

detected by WGS in the metastatic samples described in **a**.

D. 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.

E. ERα consensus peaks in primary tumors were ranked by differential signal Fold Change (Tumor vs Healthy). On the top is plotted the number of somatic variants overlapping with each individual ERα peak: higher the amount of overlapping somatic variants darker the bar color.

F. MA-plot of differential expression analyses in Ishikawa cell lines upon 8h of 10nM 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). If one these genes was differentially expressed ($|\log_2(\text{Fold Change expression})| \geq 1$ & p-value adjusted < 0.05) upon estradiol stimulation, it has been then labeled and highlighted by a red dot (n=12).

G. Ishikawa endometrial cancer cells were treated or not for 6 or 12 hours with 10nM estradiol. Whole-cell extracts were analyzed by immunoblotting with antibodies against ERα and GAPDH (loading control).

H. Progression-Free Kaplan-Meier curve of endometrial cancer patients (TCGA data) divided into two groups using the median of *ESR1* expression (FPKM) as cut-off.

Fig. 4 | Short-range chromatin contacts at the *ESR1* locus are stronger in tumors.

A. ChIP-seq genomic tracks for ERα (upper panel) and H3K27ac (lower panel) 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 40kb 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 ±500kb 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 SNPs 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.

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 ±25bp 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 cut-off.

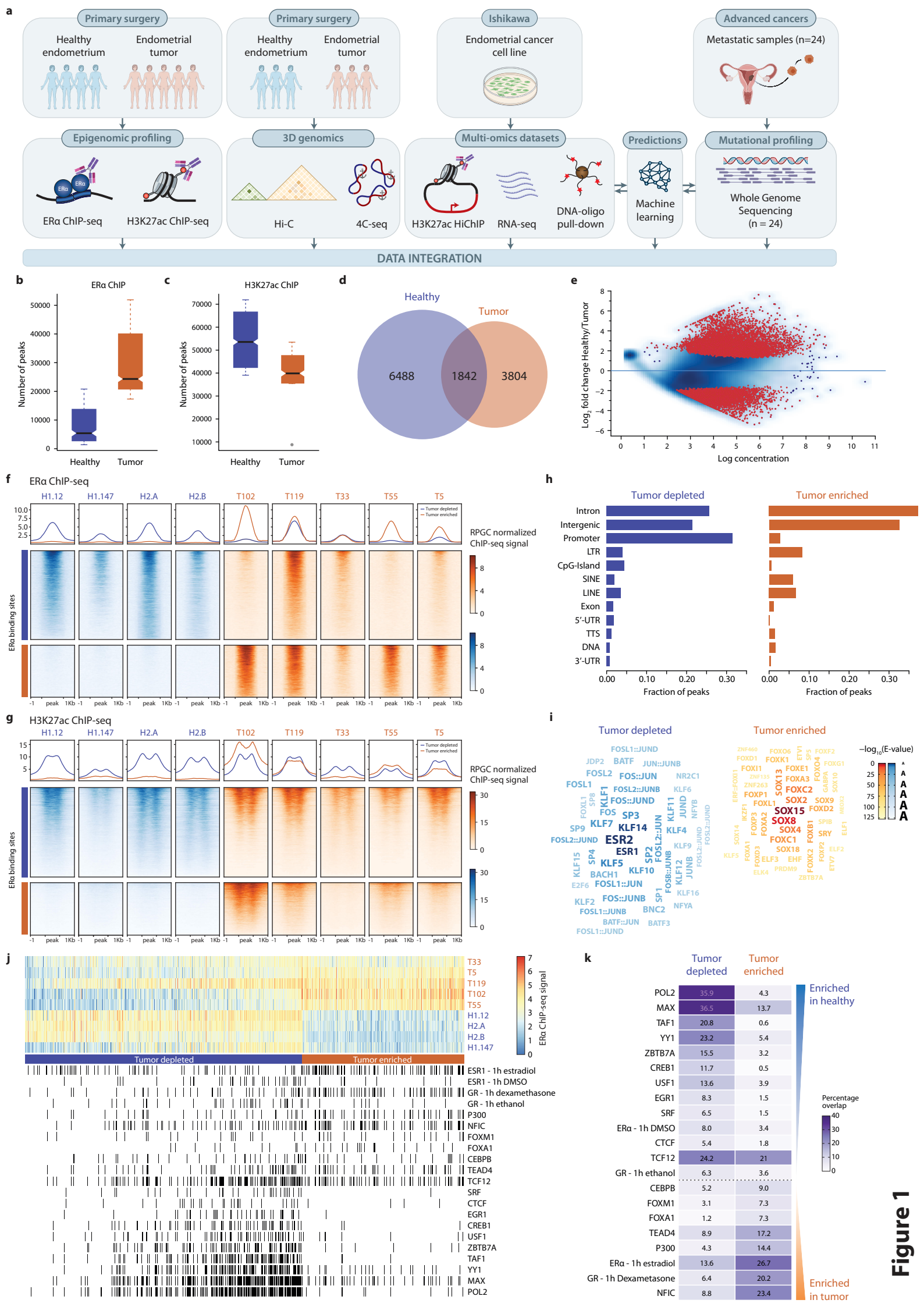
E. Ishikawa endometrial cancer cells were cross-linked and DNA sheared by sonication. Protein lysates were then immunoprecipitated (IP) with antibodies against EHMT2/G9a and IgG (negative control) to be then immunoblotted with antibodies against EHMT2/G9a and ERα. For EHMT2/G9a, two different exposure images were used for input and IP, as indicated by the vertical dashed bar.

F. Ishikawa cells were stimulated for 6h with 10nM estradiol before performing EHMT2/G9a ChIP. Quantitative PCR (qPCR) analyses were performed to determine the percentage of enrichment over the input (% Input) at the *ESR1 Enhancer 1*, *ESR1* promoter and *CDK12* promoter (negative control). Mean of 2 independent experiments is shown.

G. Ishikawa endometrial cancer cells were treated or not for 72 hours with 10nM estradiol in combination or not with 100nM ICI (Fulvestrant). 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. Schematic representation of the *ESR1* gene regulation working model. Upon tumorigenesis, ERα is re-located to an endometrial cancer specific *ESR1* enhancer where interacts with EHMT2/G9a and comes in contact with the *ESR1* promoter. These changes lead to a decrease in *ESR1* gene expression. In more advanced endometrial cancer stages, a somatic mutation is acquired at the tumor-specific *ESR1* enhancer

leading to a loss of EHMT2/G9a chromatin occupancy which allows re-expression of *ESR1*.


a
Primary surgery
Healthy endometrium
Endometrial tumor
Primary surgery
Healthy endometrium
Endometrial tumor
Ishikawa
Endometrial cancer cell line
Advanced cancers
Metastatic samples (n=24)
Epigenomic profiling
ERα ChIP-seq
H3K27ac ChIP-seq
3D genomics
Hi-C
4C-seq
Multi-omics datasets
H3K27ac HiChIP
RNA-seq
DNA-oligo pull-down
Predictions
Machine learning
Mutational profiling
Whole Genome Sequencing (n = 24)
DATA INTEGRATION
b
ERα ChIP
Number of peaks
Healthy
Tumor
c
H3K27ac ChIP
Number of peaks
Healthy
Tumor
d
Healthy
Tumor
6488
1842
3804
e
Log2 fold change Healthy/Tumor
Log concentration
f
ERα ChIP-seq
H1.12
H1.147
H2.A
H2.B
T102
T119
T33
T55
T5
Tumor depleted
Tumor enriched
RPGC normalized ChIP-seq signal
ERα binding sites
peak
1Kb
g
H3K27ac ChIP-seq
H1.12
H1.147
H2.A
H2.B
T102
T119
T33
T55
T5
Tumor depleted
Tumor enriched
RPGC normalized ChIP-seq signal
ERα binding sites
peak
1Kb
h
Tumor depleted
Tumor enriched
Intron
Intergenic
Promoter
LTR
CpG-Island
SINE
LINE
Exon
5'-UTR
TTS
DNA
3'-UTR
Fraction of peaks
Fraction of peaks
i
Tumor depleted
Tumor enriched
−log10(E-value)
j
T33
T5
T119
T102
T55
H1.12
H2.A
H2.B
H1.147
ERα ChIP-seq signal
Tumor depleted
Tumor enriched
ESR1 - 1h estradiol
ESR1 - 1h DMSO
GR - 1h dexamethasone
GR - 1h ethanol
P300
NFIC
FOXM1
FOXA1
CEBPB
TEAD4
TCF12
SRF
CTCF
EGR1
CREB1
USF1
ZBTB7A
TAF1
YY1
MAX
POL2
k
Tumor depleted
Tumor enriched
POL2 35.9 4.3
MAX 36.5 13.7
TAF1 20.8 0.6
YY1 23.2 5.4
ZBTB7A 15.5 3.2
CREB1 11.7 0.5
USF1 13.6 3.9
EGR1 8.3 1.5
SRF 6.5 1.5
ERα - 1h DMSO 8.0 3.4
CTCF 5.4 1.8
TCF12 24.2 21
GR - 1h ethanol 6.3 3.6
CEBPB 5.2 9.0
FOXM1 3.1 7.3
FOXA1 1.2 7.3
TEAD4 8.9 17.2
P300 4.3 14.4
ERα - 1h estradiol 13.6 26.7
GR - 1h Dexametasone 6.4 20.2
NFIC 8.8 23.4
Enriched in healthy
Percentage overlap
Enriched in tumor
Figure 1

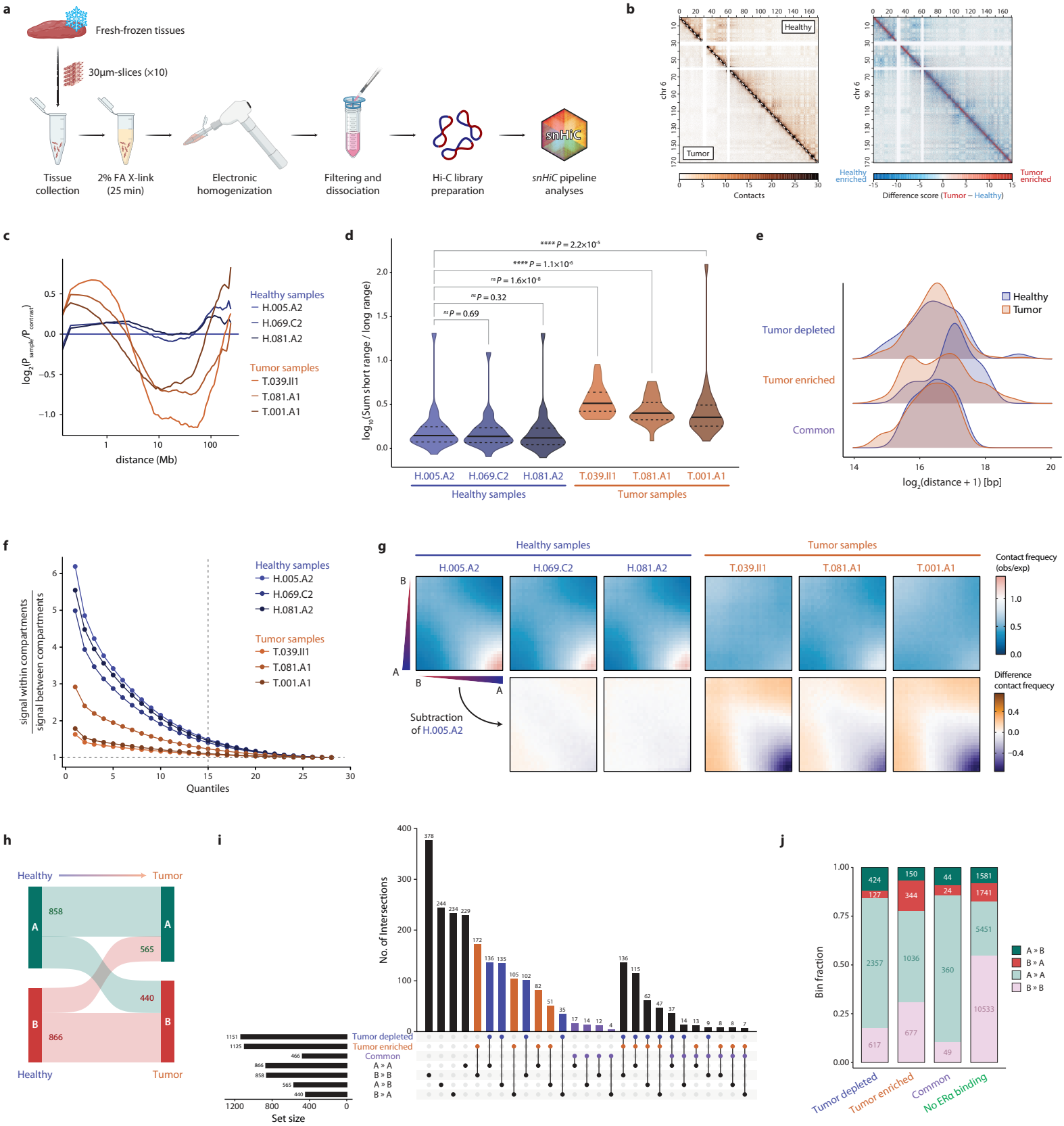

a
Fresh-frozen tissues
30µm-slices (×10)
Tissue collection
2% FA X-link (25 min)
Electronic homogenization
Filtering and dissociation
Hi-C library preparation
snHiC pipeline analyses
b
Healthy
Tumor
chr 6
Contacts
Healthy enriched
Tumor enriched
Difference score (Tumor – Healthy)
c
Healthy samples
H.005.A2
H.069.C2
H.081.A2
Tumor samples
T.039.II1
T.081.A1
T.001.A1
distance (Mb)
d
**** P = 2.2×10-5
**** P = 1.1×10-6
ns P = 1.6×10-8
ns P = 0.32
ns P = 0.69
log10(Sum short range / long range)
Healthy samples
Tumor samples
e
Healthy
Tumor
Tumor depleted
Tumor enriched
Common
log2(distance + 1) [bp]
f
signal within compartments / signal between compartments
Quantiles
g
Healthy samples
Tumor samples
Contact frequency (obs/exp)
Difference contact frequency
Subtraction of H.005.A2
h
Healthy
Tumor
858
565
440
866
i
No. of Intersections
Set size
Tumor depleted
Tumor enriched
Common
A » A
B » B
A » B
B » A
j
Bin fraction
A » B
B » A
A » A
B » B
Tumor depleted
Tumor enriched
Common
No ERα binding

Figure 2

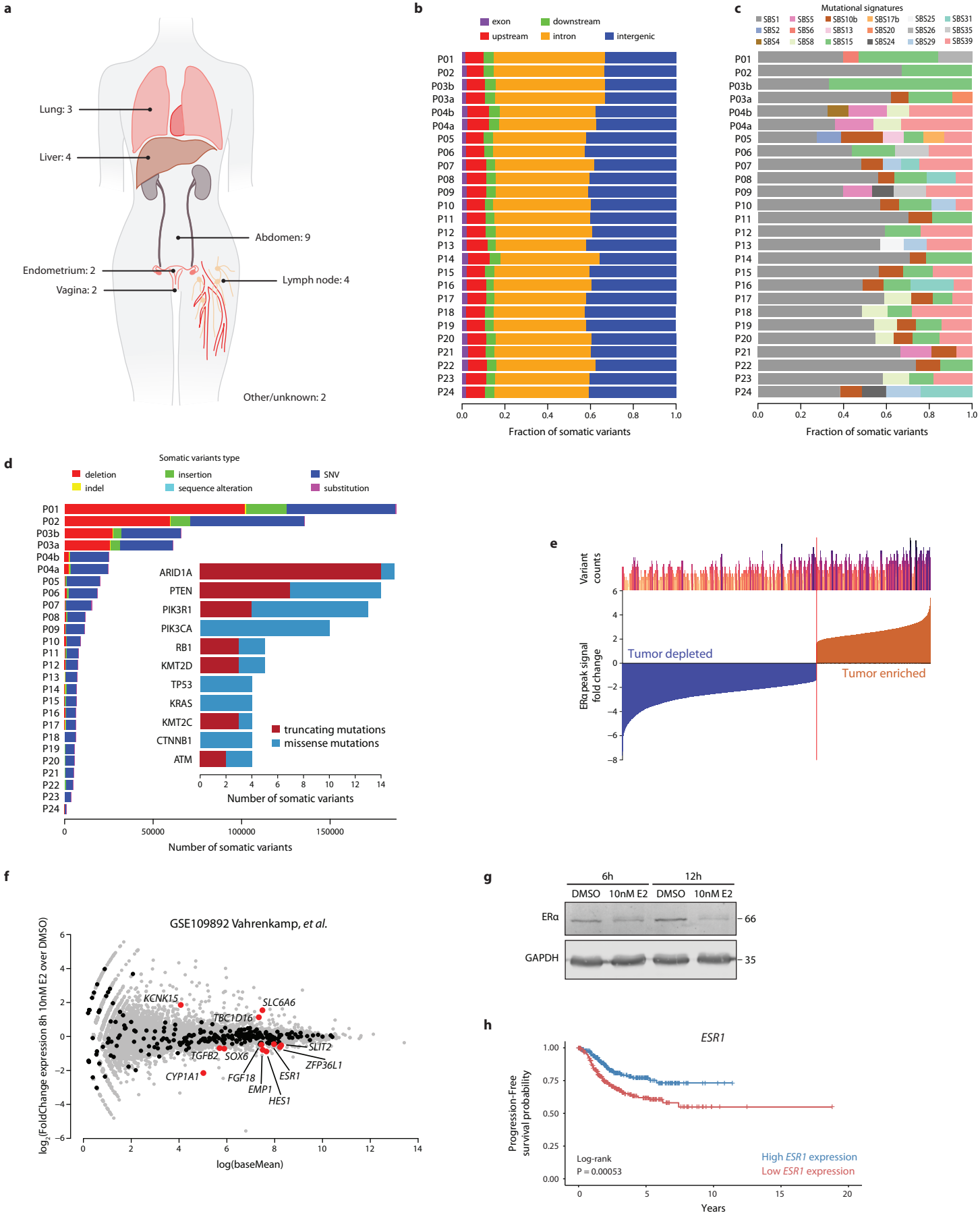

a
Lung: 3
Liver: 4
Abdomen: 9
Endometrium: 2
Vagina: 2
Lymph node: 4
Other/unknown: 2
b
exon
downstream
upstream
intron
intergenic
Fraction of somatic variants
c
Mutational signatures
SBS1
SBS5
SBS10b
SBS17b
SBS25
SBS31
SBS2
SBS6
SBS13
SBS20
SBS26
SBS35
SBS4
SBS8
SBS15
SBS24
SBS29
SBS39
Fraction of somatic variants
d
Somatic variants type
deletion
insertion
SNV
indel
sequence alteration
substitution
ARID1A
PTEN
PIK3R1
PIK3CA
RB1
KMT2D
TP53
KRAS
KMT2C
CTNNB1
ATM
truncating mutations
missense mutations
Number of somatic variants
Number of somatic variants
e
Variant counts
ERα peak signal fold change
Tumor depleted
Tumor enriched
f
GSE109892 Vahrenkamp, et al.
KCNK15
SLC6A6
TBC1D16
SLIT2
ZFP36L1
TGFB2
SOX6
CYP1A1
FGF18
ESR1
EMP1
HES1
log2(FoldChange expression 8h 10nM E2 over DMSO)
log(baseMean)
g
6h
12h
DMSO
10nM E2
DMSO
10nM E2
ERα
GAPDH
66
35
h
ESR1
Progression-Free survival probability
Log-rank
P = 0.00053
High ESR1 expression
Low ESR1 expression
Years

Figure 3

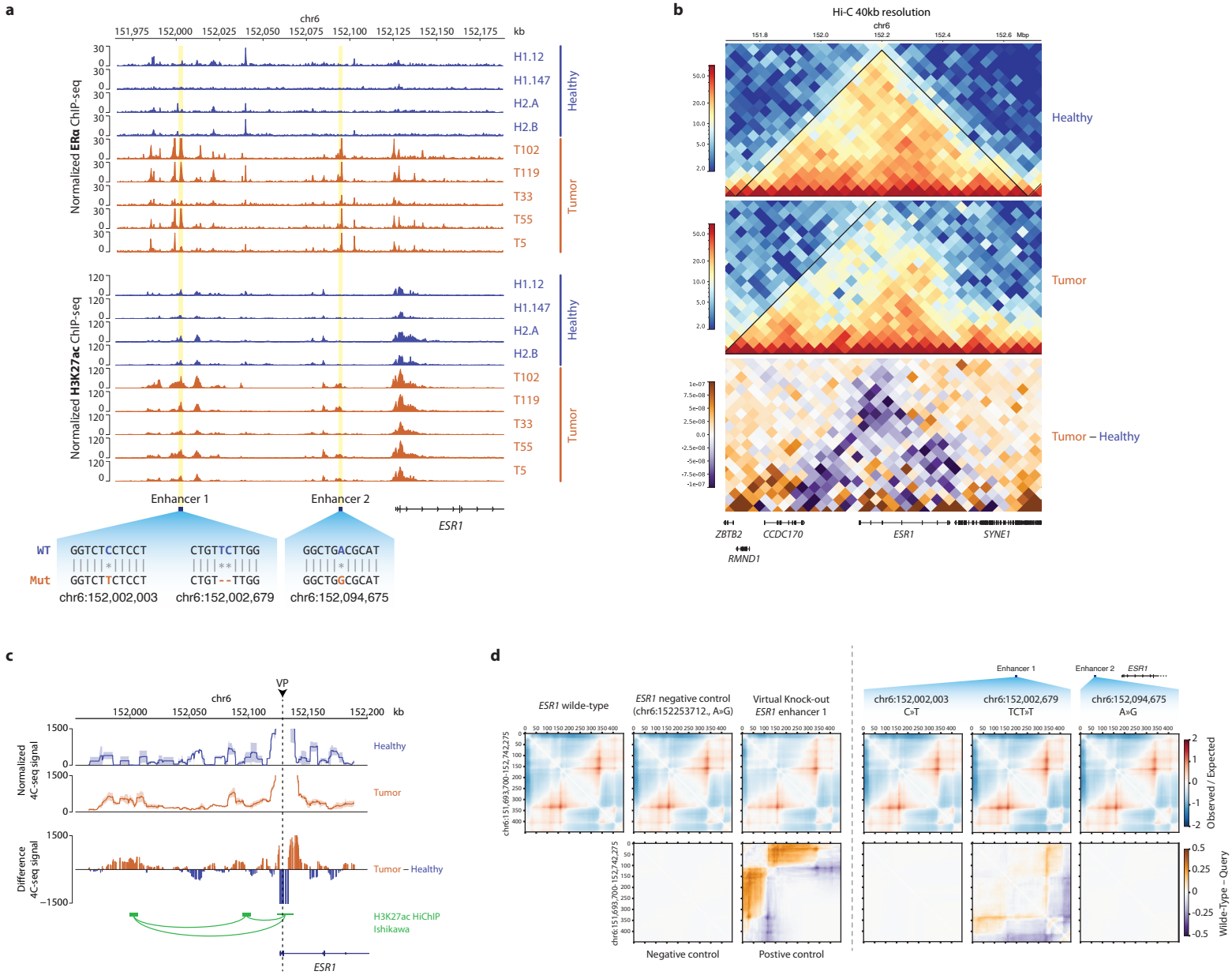

a
chr6
151,975 152,000 152,025 152,050 152,075 152,100 152,125 152,150 152,175 kb
Normalized ERα ChIP-seq
H1.12
H1.147
H2.A
H2.B
T102
T119
T33
T55
T5
Healthy
Tumor
Normalized H3K27ac ChIP-seq
Enhancer 1
Enhancer 2
ESR1
WT GGTCTCCTCCT CTGTTCTTGG GGCTGACGCAT
Mut GGTCTTCTCCT CTGT--TTGG GGCTGGCGCAT
chr6:152,002,003 chr6:152,002,679 chr6:152,094,675
b
Hi-C 40kb resolution
chr6
151.8 152.0 152.2 152.4 152.6 Mbp
Healthy
Tumor
Tumor – Healthy
ZBTB2
CCDC170
ESR1
SYNE1
RMND1
c
VP
chr6
152,000 152,050 152,100 152,150 152,200 kb
Normalized 4C-seq signal
Healthy
Tumor
Difference 4C-seq signal
Tumor – Healthy
H3K27ac HiChIP Ishikawa
ESR1
d
ESR1 wilde-type
ESR1 negative control (chr6:152253712., A»G)
Virtual Knock-out ESR1 enhancer 1
Enhancer 1
Enhancer 2
ESR1
chr6:152,002,003 C»T
chr6:152,002,679 TCT»T
chr6:152,094,675 A»G
chr6:151,693,700-152,742,275
Observed / Expected
Wilde-Type – Query
Negative control
Postive control

Figure 4

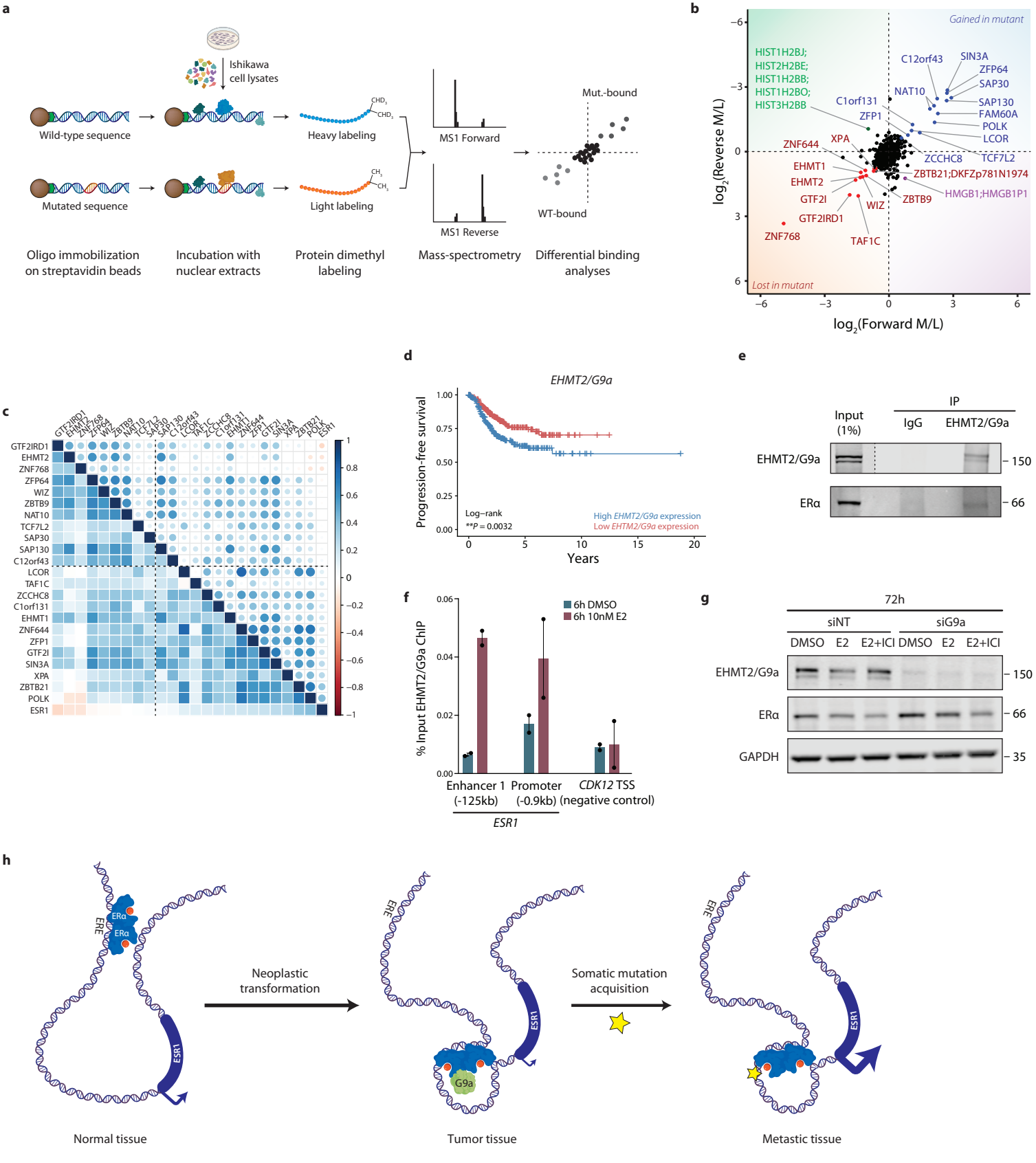

a
Ishikawa cell lysates
Wild-type sequence
Mutated sequence
Heavy labeling
Light labeling
MS1 Forward
MS1 Reverse
Mut.-bound
WT-bound
Oligo immobilization on streptavidin beads
Incubation with nuclear extracts
Protein dimethyl labeling
Mass-spectrometry
Differential binding analyses
b
Gained in mutant
Lost in mutant
HIST1H2BJ; HIST2H2BE; HIST1H2BB; HIST1H2BO; HIST3H2BB
C12orf43
SIN3A
ZFP64
SAP30
NAT10
SAP130
FAM60A
POLK
LCOR
C1orf131
ZFP1
ZNF644
XPA
ZCCHC8
TCF7L2
EHMT1
EHMT2
ZBTB21;DKFZp781N1974
HMGB1;HMGB1P1
GTF2I
WIZ
ZBTB9
GTF2IRD1
ZNF768
TAF1C
log2(Reverse M/L)
log2(Forward M/L)
c
d
EHMT2/G9a
Progression-free survival
Log–rank
**P = 0.0032
High EHMT2/G9a expression
Low EHTM2/G9a expression
Years
e
Input (1%)
IP
IgG
EHMT2/G9a
ERα
150
66
f
6h DMSO
6h 10nM E2
% Input EHMT2/G9a ChIP
Enhancer 1 (-125kb)
Promoter (-0.9kb)
CDK12 TSS (negative control)
ESR1
g
72h
siNT
siG9a
DMSO
E2
E2+ICI
EHMT2/G9a
ERα
GAPDH
150
66
35
h
ERα
ERE
ESR1
Neoplastic transformation
G9a
Somatic mutation acquisition
Normal tissue
Tumor tissue
Metastic tissue

Figure 5