ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response

DOI:
10.1038/s41588-019-0541-5

Document Version
Accepted author manuscript

1. Extended Data

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper.
Extended Data Fig. 1	Enrichment of BAF and P-BAF components in the CRISPR screen	Nagarajan ED Fig. 1.jpg	**a**. Scatterplot of CRISPR screening data, showing enrichment of BAF components following 26 days of different drug treatment, relative to DMSO treated control cells. n= 3 independent viral infections. **b**. Log2 fold changes showing gRNA enrichment/depletion against all BAF, P-BAF and ncBAF components in the CRISPR screen. Treatment conditions are compared to DMSO control. More proliferative changes represent enriched gRNA after treatment, indicating genes that contribute to drug resistance. **c, e.** Validation of ARID1A perturbation effect on proliferation and drug response using *ARID1A* siRNA on MCF7 (**c**) and ZR-75-1 (**e**), representative experiments shown from 2 similar independent experiments each cell line. p-values calculated by One way ANOVA test. * denote $p < 0.05$, *** denotes $p < 0.001$. Sample size mentioned in S4. Measure of centre represents mean ± SEM (**c**) and mean ± SD (**e**). **d**. Western blot of ARID1A protein levels after siRNA transfection in MCF7 cells. A representative image is shown from 3 similar independent experiments. Unprocessed Western blot in Source Data Fig. 2.
Extended Data Fig. 2	ARID1A co-binds ER and FOXA1-bound regulatory elements, but is depleted with	Nagarajan ED Fig.2.jpg	**a-c**. Single gene profiles showing the binding of ER, FOXA1 and ARID1A on overlapping sites in MCF7 cells. ChIP-seq was performed using three independent biological cell cultures. **d**. Overlap of binding sites for ER,

	estrogen treatment.		FOXA1 and ARID1A binding sites in ZR-75-1 cells. **e**. Boxplots showing the normalized ChIP-seq tag density around 400 bp window around the center of ARID1A binding on DiffBind-defined estrogen independent (constant) and dependent (reduced with estrogen) sites in MCF7. Both classes show reduced ARID1A binding upon estrogen. p-values were calculated by Welch's t-test, two-sided. Centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range). Statistical test details are mentioned in Supplementary Table 5e.
Extended Data Fig. 3	Enrichment of SWI/SNF factors with ER and FOXA1 in RIME	Nagarajan ED Fig.3.jpg	**a**. ARID1A and BRG1 RIME were conducted on asynchronous MCF7 cells on two biological cell cultures. Label free quantification was performed to show the log 2 scaled normalized intensities of the BAF, P-BAF, ncBAF and common subunits of SWI/SNF complex. Rabbit polyclonal IgG is used as the negative control. **b**. ER qPLEX-RIME was performed on five primary tumours from ER+ breast cancer patients and the ER interactors are shown as enrichment over IgG vs -log10 p-value, corrected by Benjamini and Hochberg multiplicity correction, two-sided. **c, d**. Boxplots illustrating the more enrichment of HDAC1 (**c**) and less enrichment of random factors (**d**) in ERα RIME in five patients compared to IgG negative control in human breast tumours. The values are scaled to the median of IgG and log2 transformed. **e.** Boxplots illustrating the enrichment of selected known ERα interactors from the RIME experiment in MCF7 cells at a representative timepoint (4-hydroxytamoxifen- 24 hrs) comparing to IgG negative control. The values are scaled to

			the median of IgG and log2 transformed. n = 5 independent biological cell cultures. For all boxplots, Centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range).
Extended Data Fig. 4	Enrichment of SWI/SNF factors during Tamoxifen and Fulvestrant in ChIP-seq experiments	Nagarajan ED Fig.4.jpg	**a-d.** Asynchronous MCF7 cells were treated with vehicle or Fulvestrant, an ER degrader and ChIP-seq was conducted for ARID1A (**b**), BRG1 (**c**) or SNF5 (**d**). Triplicate independent cell cultures were conducted. **d**. Single gene profile showing the induction of SWI/SNF complex binding during Fulvestrant treatment. **e**. Overlap of ARID1A lost sites during estrogen treatment with gained sites during Tamoxifen and Fulvestrant from three independent biological cell cultures. **f**. Overlap of ARID1A gained sites during Tamoxifen treatment with Fulvestrant and Tamoxifen downregulated genes.
Extended Data Fig. 5	FOXA1 promotes the binding of ARID1A and BRG1.	Nagarajan ED Fig.5.jpg	Hormone-deprived ZR-75-1 cells were transfected with control or *FOXA1* siRNA and ChIP-seq was conducted for ARID1A (**a**) and BRG1 (**b**). n = 3 independent biological cell cultures. MA plots are shown with the average intensity of binding vs log2 fold change with *FOXA1* siRNA relative to control siRNA. **c**. Scatterplot showing the association of the loss of ARID1A and BRG1 binding upon *FOXA1* knockdown. PCC – Pearson Correlation coefficient, two-sided. **d**. Heatmaps shown on ARID1A and BRG1 FOXA1 independent (common) and dependent (lost sites with *FOXA1* knockdown) sites in ZR-75-1 cells. **e**. Boxplots showing the normalized ChIP-seq tag density around 400 bp window of ARID1A and BRG1 on FOXA1 independent (constant, n=70,429 sites) and dependent (lost sites with si*FOXA1*, n=17,357 sites) sites in ZR-75-1. p-value calculated by Welch's

			test, two-sided. n = 3 independent biological cell culture samples. Centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range). Statistical test details are mentioned in Supplementary Table 5f.
Extended Data Fig. 6	FOXA1 promotes the binding of ARID1A and BRG1.	Nagarajan ED Fig.6.jpg	Hormone-deprived MCF7 and ZR-75-1 cells were transfected with control or *FOXA1* siRNA and ChIP-seq was conducted for ARID1A and BRG1. n = 3 independent biological cell cultures. (**a-b**) Single gene profiles of *CCND1* (**a**) and *CDH1* (**b**) showing the effect on SWI/SNF complex binding with *FOXA1* knockdown on MCF7 and ZR-75-1 cells. ER and FOXA1 binding overlap is shown. (**c-d**) ChIP-qPCR analyses on specific sites (*CCND1* and *CDH1* ER binding sites) showing ARID1A and BRG1 binding with *FOXA1* knockdown in hormone-deprived MCF7 and ZR-75-1 cells (**c**) or ARID1A binding following Tamoxifen treatment in asynchronous MCF7 cells (**d**). n = 3 independent biological cell cultures. * denotes $p \leq 0.05$, ** denotes $p \leq 0.01$, *** denotes $p \leq 0.001$. Precise p-values are mentioned in Fig. S10. Mean is measured as centre shown with standard deviation. Details of the statistical tests are mentioned in Fig. S10.
Extended Data Fig. 7	ATAC-seq analyses shows a negligible regulation of ARID1A on transcription-associated chromatin opening.	Nagarajan ED Fig.7.jpg	**a.** Heatmap showing ATAC-seq analysis in *ARID1A* KO clones 11 and 14 following Tamoxifen treatment. Common, gained and lost sites defined by DiffBind analysis. n = 4 independent biological cell cultures. FDR ≤ 0.05 corrected by Benjamini-Hochberg multiplicity correction, two-sided. **b.** Association of *ARID1A* KO upregulated and downregulated genes with ATAC-seq gained and lost sites.

Extended Data Fig. 8	ARID1A perturbation regulates ARID2 binding.	Nagarajan ED Fig.8.jpg	**a**. ARID2 ChIP-seq was conducted in wild type cells or the two *ARID1A* knock-out clonal cell lines and heatmaps are shown on ARID2 binding sites after Tamoxifen treatment. Also included was ARID1A ChIP-seq from wild type cells treated with vehicle or Tamoxifen. ARID2 binding overlapped with ARID1A binding and was dependent on ARID1A. n = 3 independent biological cell cultures. **b.** Signal intensity plot showing changes in ARID2 binding in wild type control cells or ARID1A knock-out cells at ARID2 binding sites. n = 3 independent biological cell cultures.
Extended Data Fig. 9	ARID1A promotes BRG1 and HDAC1 binding without affecting ER and H3K27ac occupancy	Nagarajan ED Fig.9.jpg	**a, b**. BRG1, H3K27Ac, HDAC1 and ER (**b**) ChIP-seq were conducted in asynchronous wild type cells treated with vehicle or tamoxifen or in the two *ARID1A* knock-out clones (Clones 11 and 14) following tamoxifen treatment. The binding is shown on regions where HDAC1 is lost in *ARID1A* knockout cells relative to wild type cells. n = 3 independent biological cell cultures. **c, d**. Scatterplot showing the correlation of ER (**c**) or H3K27Ac (**d**) and HDAC1 binding in *ARID1A* knockout clone 11 versus wild type cells. n = 3 independent biological cell cultures. PCC – Pearson Correlation coefficient. p-values were calculated by Pearson correlation test, two-sided. **e**. Principal Component Analysis (PCA) of normalised peptide intensities of PDX tumours after ER qPLEX-RIME. n= 2 PDX each group. **f**. Details of *ARID1A* mutations observed within ER+ PDX tumours used in ER qPLEX-RIME.
Extended Data Fig. 10	ARID1A regulates histone H4 acetylation.	Nagarajan ED Fig.10.jpg	Upregulation of histone H4 acetylation in *ARID1A* knock-out clone 11 and 14 in Vehicle (**a**) or Tamoxifen (**b**) treated cells comparing to wild type cells. Heatmap representing the changes in histone H4Ac marks upon *ARID1A* knockout with Vehicle or Tamoxifen treatment on ER binding sites close to

			ARID1A repressed genes. n =3. (**c**) Empirical cumulative probability distribution plots of H4K8Ac and H4K12Ac ChIP-seq signals showing upregulation in intensity (y-axis) with *ARID1A* knockouts clones 11 and 14. Plots were made on ER sites close to ARID1A repressed genes (n=686 sites) with more than 75% contribution to the variance in intensity. Window – 2 kb around the center of binding.

2. Supplementary Information:

A. Flat Files

Item	**Present?**	**Filename** **This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf**	**A brief, numerical description of file contents.** **i.e.: *Supplementary Figures 1-4, Supplementary Note, and Supplementary Tables 1-4.***
Supplementary Information	Yes	Supplementary Information.pdf	Supplementary Figures 1-15 and Supplementary Note.
Reporting Summary	Yes	Reportingsummary.pdf	

B. Additional Supplementary Files

Type	**Number** **If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.**	**Filename** **This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_ Supplementary Video 1.mov***	**Legend or Descriptive Caption** **Describe the contents of the file**
Supplementary Table	1	Supplementary Table 1.xlsx	Supplementary Tables 1-6.

3. Source Data

Figure	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_Source Data Fig1.xls,* or *Smith_ Unmodified Gels_Fig1.pdf*	Data description i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1	Nagarajan Source Data Fig.1.pdf	Unprocessed Western Blots for Extended Data Fig. 1d.
Source Data Fig. 2	Nagarajan Source Data Fig.2.pdf	Unprocessed Western Blots for Fig. 2a.

ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response

Sankari Nagarajan[1], Shalini V. Rao†[1,2], Joseph Sutton†[1], Danya Cheeseman†[1], Shanade Dunn[3], Evangelia K. Papachristou[1], Jose-Enrique Gonzalez Prada[1], Dominique-Laurent Couturier[1], Sanjeev Kumar[1], Kamal Kishore[1], Chandra Sekhar Reddy Chilamakuri[1], Silvia-Elena Glont[1], Emily Archer Goode[1], Cara Brodie[1], Naomi Guppy[4], Rachael Natrajan[4], Alejandra Bruna[1], Carlos Caldas[1], Alasdair Russell[1], Rasmus Siersbæk[1], Kosuke Yusa[3,5], Igor Chernukhin[1] and Jason S. Carroll[1]*

[1]CRUK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge, CB2 0RE, UK.

[2]Dept of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

[3]Wellcome Trust Sanger Institute, Hinxton, UK.

[4]The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London.

[5]Institute for Frontier Life and Medical Sciences, Kyoto University, Japan,

*Correspondence to: Jason.carroll@cruk.cam.ac.uk.

† Equal contribution

Abstract

Using genome-wide CRISPR screens to understand endocrine drug resistance, we discovered *ARID1A* and other SWI/SNF complex components as the most critical factors required for response to two classes of Estrogen Receptor-alpha (ER) antagonists as these SWI/SNF-specific gene knockouts lead to drug resistance. Unexpectedly, *ARID1A* was also the top candidate for response to the BET inhibitor JQ1, but in the opposite direction, where loss of *ARID1A* sensitised breast cancer cells to BET inhibition. We show that ARID1A is a repressor which binds chromatin at ER cis-regulatory elements. However, ARID1A elicits repressive activity in an enhancer-specific, but FOXA1-dependent and active ER-independent manner. Deletion of ARID1A resulted in loss of Histone Deacetylase 1 (HDAC1) binding, increased histone 4 lysine acetylation and subsequent BRD4-driven transcription and growth. *ARID1A* mutations are more frequent in treatment-resistant disease and our findings provide mechanistic insight into this process whilst revealing rational treatment strategies for these patients.

Key words: ARID1A, Breast cancer, Treatment resistance, CRISPR screens

Introduction

Three quarters of breast cancers are driven by Estrogen Receptor-alpha (ER) [1], which utilises a slew of associated proteins to access compacted chromatin (including Forkhead Box A1 (FOXA1) and GATA Binding Protein-3 (GATA3)) [2, 3]. Drugs that target the ER pathway are effective treatments for a majority of women with ER+ disease [1], but a substantial fraction of women will present with *de novo* or acquired drug resistance. Mechanisms of resistance are varied and include changes in co-factor levels, growth factor activated transcription and mutations in ER and associated transcription factors and co-factors [4].

Significant effect has been invested in identifying associated protein complexes that influence ER transcriptional activity [5-7]. A role for the ATP-dependent chromatin remodeling complex SWItch mating type/Sucrose Non-Fermenting chromatin remodeling complex (SWI/SNF), has been linked with nuclear receptor function [8,9], where this complex modulates chromatin accessibility. There are three ATPase complexes, BAF, P-BAF and a recently identified non-canonical BAF (ncBAF) and the BRG1 and BRM subunits are common between the three complexes. However, there are proteins that are specific to BAF (ARID1A, ARID1B, DPF1/2/3, SS18), P-BAF complex (ARID2, Polybromo (PBRM1), BRD7) and ncBAF (BRD9, GLTSCR1, GLTSCR1L) [10,11]. Previous work has shown a physical association between the SWI/SNF component BRG1 and ER and a requirement for BRG1 for ER-mediated transcriptional activity [12,13]. The recruitment of SWI/SNF to the ER complex, is mediated by shared co-factors [14] and BRG1 occupancy at ER regulatory elements, coincides with increased localised histone acetylation [15]. On a locus-specific level, BRG1 can bind to ER regulatory elements independent of ER [12], suggesting that the SWI/SNF complex might contribute to chromatin preparation prior to ER recruitment.

The SWI/SNF complex is important for chromatin regulation and gene expression [16], it is mutated in ~20% of all human cancers [17] and has been linked with the transcriptional activity of numerous nuclear receptors [8,9,13,18]. Wild type *ARID1A* expression is associated with better

clinical outcome in ER+ breast cancer patients [19] and importantly *ARID1A* inactivating mutations are enriched in treatment-resistant tumours and metastases (in total 12% of cases) [20,21]. In addition, *ARID1A* inactivation has been associated as a tumour promoting event in ER+ breast cancer [22].

To systematically identify genes involved in treatment response in breast cancer, we employed global Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) screening approaches, coupled with three different treatment modalities, which revealed a role for the SWI/SNF complex, as critical determinants of treatment response.

Results

A CRISPR screen reveals ARID1A as a gene involved in treatment response: We employed a CRISPR screening approach, which encompassed gRNAs that target a total of 18,009 human genes [23]. We established Cas9-expressing MCF7 breast cancer cells (Supplementary Fig. 1) which were infected and grown for 20 days. All cell line experiments were conducted in asynchronous cells grown in estrogen-rich media. Three biological cell cultures with independent viral infections with CRISPR vectors were performed as described in the Online Methods section. Analysis of the depleted gRNAs at different post-infection time-points, revealed known ER interactors including Cyclin D1 (*CCND1*), *FOXA1* and *GATA3* (Fig. 1a and Supplementary Fig. 2, Supplementary Table 1), albeit with different essentiality kinetics (Fig. 1a and 1b). In addition, a number of gRNAs were enriched representing tumour suppressors or growth inhibitors (Fig. 1c). As expected, growth promoting genes required for cellular viability showed greater gRNA depletion with longer infection (Fig. 1d). After 9 days of infection, we subsequently treated cells for a total of 26 days with the Selective Estrogen Receptor Modulator (SERM) 4-hydroxytamoxifen (Tamoxifen) or the Selective Estrogen Receptor Degrader (SERD) Fulvestrant (ICI 182780). We also used the tool compound JQ1, which targets Bromodomain and Extraterminal Domain (BET)-containing proteins, since

Bromodomain containing protein-4 (BRD4) is postulated to be a therapeutic target in ER+ breast cancers and BET inhibitors are currently being explored in clinical trials [24,25]. Three independent infections were performed and the data was integrated as described in the methods sections (Complete data in Supplementary Table 2). When specifically assessing genes required for treatment response, we found that the Fulvestrant and Tamoxifen CRISPR screens looked largely similar (Fig. 1e). Despite the distinct mechanisms of growth suppression (Fulvestrant degrades ER, whereas Tamoxifen-bound ER is recruited to the chromatin as a repressive complex), 63.5% of the genes required for Fulvestrant's antiproliferative effects were also required for Tamoxifen activity (Fig. 1e and Supplementary Fig. 2). One of the most significantly enriched gene was AT-Rich Interaction Domain 1A (*ARID1A)*, a component of the BAF ATP-dependent chromatin remodeling complex. It was one of the most essential gene for both Tamoxifen and Fulvestrant activity and depletion of *ARID1A* (i.e. enrichment of gRNAs targeting *ARID1A*) resulted in drug resistance to both compounds. Unexpectedly, *ARID1A* was the highest ranked gene in the JQ1 treated cells (ranked 1 out of 18,009 genes), but in the opposite direction, where gRNAs were observed to be depleted in JQ1 treated conditions (Fig. 1f, 1g and Supplementary Fig. 2). Other BAF components, including *ARID1B*, SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily B, Member-1 (*SMARCB1*/BAF47/SNF5) and Synovial Sarcoma Translocation, Chromosome 18 (*SS18*) showed the same pattern (Fig. 1f, 1g and Extended Data 1), suggesting that the BAF complex is required for ER targeted drugs to work, but when lost, sensitises cells to BET inhibitors. The dependence on *ARID1A* for growth arrest mediated by ER-targeted agents was validated in MCF7 and ZR-75-1 cells using *ARID1A* siRNA (Extended Data Fig. 1, Source Data Fig. 1 and Supplementary Fig. 3).

Genomic characterisation of ARID1A function: We subsequently assessed the potential genomic interplay between ARID1A and ER. We performed three independent biological

replicates of ChIP-seq for ARID1A in MCF7 and ZR-75-1 cells and peaks were called using MACS version 2 [26], resulting in 21,226 ARID1A peaks in MCF7 and 56,966 peaks in ZR-75-1. ARID1A binding sites were found to commonly co-occur at ER and FOXA1 binding events (Fig. 1h and Extended Data 2) and global analysis revealed that more than 78% of all ARID1A binding events were shared with ER, FOXA1 or both proteins in MCF7 (Fig. 1i), implying a functional connection between ARID1A and the regulatory elements occupied by the ER/FOXA1 complex. Interestingly, ARID1A overlapped more with FOXA1 (78% ARID1A binding sites were co-bound by FOXA1) than with ER (66%) in ZR-75-1 cells (Extended Data 2). We assessed whether ARID1A binding to ER bound enhancers was dependent on ER, by hormone depriving cells, treating with vehicle (ethanol) or estrogen for 6hr and conducting ChIP-seq. ARID1A was able to bind to ER/FOXA1 binding events prior to ligand induced ER recruitment (Fig. 1j and Extended Data 2). These findings suggest that ARID1A is not a classic ER-associated co-factor and can bind to regulatory elements independent of active ER, likely in a repressive manner.

To validate the CRISPR screen, we specifically deleted *ARID1A* from MCF7 cells, resulting in two separate *ARID1A* knock-out clones (Clones 11 and 14). *ARID1A* deletion was confirmed by Sanger and amplicon-based next generation sequencing and Western blotting (Fig. 2a, Supplementary Fig. 4, Source Data Fig. 2) and potential off-target effects were assessed. *In vitro* growth of these clones and the wild type control (WT clone 219) validated the CRISPR screening results, showing that both clones had increased intrinsic proliferation and were resistant to Tamoxifen, but showed sensitivity to JQ1 (Fig. 2b and Supplementary Fig. 5) and two additional clinically relevant BET inhibitors, OTX015 (from OncoEthix/Merck) and IBET762 (from GlaxoSmithKline) (Supplementary Fig. 4).

We established xenograft tumours from the wild type or the two *ARID1A* knock-out clones in the presence of estrogen pellets to maintain ER+ tumour growth and subsequently treated cells with vehicle or 4-hydroxytamoxifen. Tumour growth at day 25 was increased in the two *ARID1A*

knock-out clones in the presence of 4-hydroxytamoxifen, when compared to wild type mice (Supplementary Fig. 5 which includes details of the statistical tests), validating that *ARID1A* is required for antiestrogen efficacy. However, the greatest difference in growth rate was in *ARID1A* wild type versus knock-out contexts in non-treated conditions (Fig. 2c and Supplementary Fig. 5) and we postulated that the diminution in Tamoxifen efficacy in *ARID1A*-null tumours may simply be due to an increased overall intrinsic proliferative potential.

ARID1A regulates ER target genes and is part of the ER complex: To explore the mechanistic role of *ARID1A* in drug response, RNA-seq was conducted using four biological cell culture samples of the wild type or *ARID1A* knock-out lines, treated with vehicle, Fulvestrant, 4-hydroxytamoxifen or BETi (JQ1). Gene expression analysis of the *ARID1A* knock-out clones and controls revealed several findings. The control lines looked similar, regardless of whether they were parental cells or wild type clonal lines (Supplementary Fig. 8). Whilst Fulvestrant and Tamoxifen showed similar gene repression patterns, JQ1 treatment resulted in a substantially different gene expression profile (Fig. 2d and Supplementary Fig. 8). In the *ARID1A* knock-out clones, JQ1 treatment showed a more consistent expression pattern when compared to the wild type cells, whereas the majority of genes repressed by Fulvestrant/Tamoxifen, were up-regulated or not changed in the *ARID1A* knock-out cells (Fig. 2d and Supplementary Fig. 6). In total, 86% of the Fulvestrant and 85% of the Tamoxifen-repressed genes were no longer significantly repressed in the *ARID1A* knock-out cells and a cluster of them (highlighted in Fig. 2d) are significantly downregulated by JQ1 treatment, to the same degree as in wild type cells. *ARID1A* deletion therefore, resulted in induction of the Fulvestrant/Tamoxifen repressed genes, even in the absence of an ER antagonist, implying ARID1A-mediated basal repression of the ER target genes. We generated a gene signature from the RNA-seq data and could show in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort of ER+ breast cancer patients [27] that the *ARID1A* repressed genes in both vehicle and anti-estrogen

conditions (those that were up-regulated in the *ARID1A* knock-out cell lines) were associated with poor clinical outcome when up-regulated in patients (Fig. 2e and Supplementary Fig. 6 which includes details of the statistical tests), again supporting the notion that ARID1A can repress genes linked with clinical outcome.

To understand the mechanism behind ARID1A regulation of innate proliferation, we used an unbiased proteomic approach called RIME (Rapid IP-Mass Spec of endogenous interactions) combined with a label-free quantification method [28] to identity interactors of ARID1A, BRG1 or ER, from asynchronous MCF-7 cells, using an IgG pulldown as a negative control (information is provided in supplemental material) (Supplementary Fig. 7). ARID1A and BRG1 purification revealed almost all the known BAF components, as well as ER and similarly, the ER RIME contained ARID1A and BRG1 in the complex (Fig. 3a, Extended Data 3 and Supplementary Table 3 and 4). The other ATP-ase complexes, P-BAF and ncBAF [11], are identified in BRG1 pulldown, but not in the ARID1A pulldown. BRG1 RIME identified all the BAF, P-BAF and ncBAF components, validating that BRG1 is common to these complexes [21]. It also showed enrichment of GLTSCR1/GLTSCR1L (BICRA/BICRL) subunits. BRG1 RIME revealed BET proteins as interactors (data not shown). We extended on these observations by re-analysing our recently published ER quantitative multiplexed RIME (qPLEX-RIME) data from five ER+ primary tumour samples from different patients [29]. We discovered ARID1A and several SWI/SNF components, including BRG1, BRM, BAF57, BAF170 and BAF155 as physical interactors of ER, even in surgical tumour tissue (Fig. 3b, Extended Data 3). Importantly, we also observed an interaction between ER and BRD4, a target of the BETi, in the patient tumour material (Fig. 3b), verifying physical associations between endogenous ER, the SWI/SNF complex and BRD4 *in vivo*. We re-analysed our previous proteomic data [29] to identify proteins that interact with Tamoxifen-bound ER [29]. ARID1A, BRG1 and a number of additional SWI/SNF components were enriched with Tamoxifen-liganded ER complex after treatment with 4-

hydroxytamoxifen for 6hr (Fig. 3c and Extended Data 3), confirming that the SWI/SNF-ER complex formation is repressive.

To explore the putative functional connection between SWI/SNF and the ER complex, we conducted a series of ChIP-seq experiments to map binding sites for ARID1A and two SWI/SNF common proteins, BRG1 and SNF5 (BAF47), in estrogen-rich asynchronous MCF7 cells treated with control or 4-hydroxytamoxifen for 6hr. Three independent biological replicates were conducted. Binding of all three proteins were increased globally following 4-hydroxytamoxifen treatment (Fig. 3d and Supplementary Fig. 11), supporting the hypothesis that they were involved in drug responsiveness. Both induced BRG1 and SNF5 sites overlapped with induced ARID1A sites, ER and FOXA1 (Fig. 3d and Supplementary Fig. 8). Unexpectedly, binding of these proteins were also increased following Fulvestrant treatment (Extended Data 4). The Fulvestrant-induced sites overlapped with both the Tamoxifen gained sites and estrogen lost sites from Fig. 1j (Extended Data 4), implying that these are the consistent hormone-regulated SWI/SNF binding regions. Altogether, our findings suggest that the recruitment of these factors, whilst able to associate with the ER complex, can bind to chromatin in an ER independent manner, in support of data showing basal repression of ER target genes by the BAF complex (Fig. 2d and 2e).

FOXA1 recruits ARID1A to chromatin: As Fulvestrant and Tamoxifen both increased BAF binding to chromatin, we speculated that the pioneer factor FOXA1, might modulate ARID1A and BRG1 recruitment to the chromatin, as supported by the data showing considerable overlap between ARID1A and FOXA1 binding (Fig. 1i and Extended Data 2). MCF7 and ZR-75-1 cells were hormone-deprived and transfected with *FOXA1* or control siRNA and ChIP-seq of ARID1A or BRG1 was conducted. Both ARID1A and BRG1 binding was substantially reduced following *FOXA1* silencing at enhancers, in both the cell lines assessed (Fig. 4a-e, Extended Data 5-6 and Supplementary Fig. 9-10), suggesting a degree of dependence on the pioneer factor

FOXA1 for SWI/SNF recruitment. Importantly, the *FOXA1*-dependent ARID1A binding sites were the same regions where Tamoxifen induced ARID1A binding to the genome (Fig. 4f-g). To understand the importance of FOXA1 on ARID1A dependent genes, we identified the ER bound *cis*-regulatory elements close to ARID1A-repressed genes (those up-regulated in *ARID1A* knock-out cells), which we had previously shown to correlate with clinical outcome (Fig. 2e). We observed a modest change on ARID1A and BRG1 recruitment on these sites with *FOXA1* loss (Fig. 4h). These findings show that the key ARID1A binding events are mediated by FOXA1 and not ER.

We sought to identify the molecular mechanism that dictated decreased drug responsiveness when SWI/SNF components were deleted (Fig. 1f and Extended Data 1). We performed Assay for Transposase-Accessible Chromatin (ATAC)-sequencing on MCF7 *ARID1A* knockout or wild type control cells, to assess if ARID1A was required for maintaining chromatin accessibility. Four independent cell culture samples were performed. We observed 233,862 total accessible regions in the genome, of which 83% (n=194,341) were not altered in *ARID1A* knock-out cells. Only 0.7% of sites showed a gain in accessibility in *ARID1A* knock-out cells and 16.3% of sites (n=38,002) sites had decreased accessibility in *ARID1A* knock-out cells (Extended Data 7). Integrative analyses of the chromatin accessibility and gene expression datasets showed that genes which are up-regulated in *ARID1A* knock-out cells are more associated with the ATAC-seq gained sites, implicating ARID1A in basal repression of these targets via inhibition of chromatin accessibility (Extended Data 7). However, there was no significant difference in accessibility at the regions co-bound by ARID1A and ER (data not shown), suggesting that loss of ARID1A is not altering chromatin accessibility at the regulatory regions bound by these protein complexes.

As previous work showed that a SWI/SNF subunit BRD9 inhibition results in a switch to P-BAF activity [18], we hypothesised that loss of ARID1A and BAF activity might result in a switch to a P-

BAF-driven pathway. We therefore conducted ChIP-seq of ARID2 (a P-BAF-specific complex component) and BRG1 in wild type or *ARID1A* knock-out clonal cell lines and could show that ARID2 binding was not appreciably changed by Tamoxifen treatment and there was substantially less ARID2 binding in both *ARID1A* knock-out clones, regardless of the hormonal treatment conditions (Extended Data 8). This is a possible consequence of the decreased overall BRG1 binding in the *ARID1A* deleted cells (Fig. 5a and Extended Data 9). As such, loss of ARID1A does not result in recruitment of ARID2 and a switch to P-BAF dependency.

ARID1A contributes to HDAC1 recruitment and mediating acetylation: To assess the mechanistic basis for the ARID1A repressive function, we performed H3K27Ac ChIP-seq and found that it was not affected in the *ARID1A* knock-out versus wild type cells (Extended Data 9). To identify other possibilities explaining the sustained gene expression in the presence of ER targeted drugs, when *ARID1A* was suppressed, we explored our RIME data and found that the histone deacetylase protein HDAC1 was an ARID1A interacting protein in non-treated conditions (Fig. 3a). In addition, in our qPLEX-RIME data, HDAC1 recruitment to the ER complex was enriched following Tamoxifen treatment, during active gene repression [29]. Furthermore, HDAC1 was one of the most statistically enriched ER interactors in ER+ primary tumour samples (Extended Data 3) compared to IgG controls. We therefore conducted HDAC1 ChIP-seq and found a substantial decrease in HDAC1 recruitment, when *ARID1A* was specifically knocked-out (Fig. 5b and Extended Data 9). Only modest changes in ER binding were observed on HDAC1 lost sites (Fig. 5c, Extended Data 9 and Supplementary Fig. 11). We also observed a modest decrease in global FOXA1 binding as determined by ChIP-seq (data not shown). However, this was explained by a parallel decrease in *FOXA1* expression, suggesting that ARID1A does not directly modulate FOXA1 recruitment to *cis*-regulatory elements and moderately influences FOXA1 binding by affecting total levels of this pioneer factor. BRG1 and HDAC1 binding was decreased at the same genomic regions in both the

ARID1A knock-out clones (Fig. 5d and 5e), suggesting that both HDAC1 and BRG1 binding was dependent on ARID1A.

Additionally, we performed ER qPLEX-RIME on four ER+ Patient-Derived Xenograft (PDX) tumours [30], including two that had *ARID1A* loss via mutation and two *ARID1A* wild type control models (Extended Data 9 and Supplementary Fig. 12). We found a decrease in interactions between ER and HDAC1, BAF170 and BAF155 in the *ARID1A* mutant PDX models (Fig. 5f). As such, BRG1-associated SWI/SNF complex physically associates with HDAC1 in an *ARID1A*-dependent manner and the transcriptional repression elicited by HDAC1 requires functional ARID1A.

Given that HDAC proteins can actively remove the acetylation marks that are read by BET proteins [31], we speculated that changes in HDAC activity might explain the increased sensitivity to BET inhibition in the absence of a functional SWI/SNF complex. The histone acetylation marks that are read by BET proteins include Histone 4 lysine residues, including H4K5Ac, H4K8Ac and H4K12Ac [31]. We assessed for increases in these histone marks in our *ARID1A* knock-out cells as a potential consequence of decreased HDAC1 recruitment. A distinct subset of histone H4 acetylated sites were increased under both non-treated and Tamoxifen treated conditions in the *ARID1A* knock-out cells, with the most prominent change observed in H4K8Ac (Supplementary Fig. 11). To understand the function of H4Ac upon ARID1A dependent genes, we examined the adjacent ER bound *cis*-regulatory elements on ARID1A target genes in *ARID1A* wild type versus knock-out cells. ARID1A was recruited to these enhancers in wild type cells and these sites showed substantial upregulation of the histone H4 acetylation, particularly H4K8/12Ac in both the clones (Fig. 5g and Extended Data 10). Given the decreased HDAC1 recruitment, the increase in H4K8/12Ac in *ARID1A* depleted cells and the increased responsiveness to BETi in *ARID1A* deleted contexts (Fig. 1f and 1g), we hypothesised that depletion of *ARID1A* would result in gained BRD4 binding and activity. BRD4 ChIP-seq in wild

type and *ARID1A* knock-out cells, revealed a gain of 6,197 BRD4 binding sites in *ARID1A* depleted cells, confirming a significant increase in BRD4 chromatin binding. Analyses on ER binding sites close to ARID1A target genes showed increased BRD4 binding under *ARID1A* loss in both treatment conditions (Fig. 6a and 6c). The same regions showed a gain of histone H4 acetylation and BRD4 and decreased HDAC1 binding in *ARID1A*-deleted cells (Fig 6b and Supplementary Fig. 13-14). We integrated the gained BRD4 binding that was only observed in *ARID1A*-null cells, with the Fulvestrant/Tamoxifen-repressed genes and found a significant enrichment of BRD4 recruitment to the genes typically repressed by both ER antagonists (Supplementary Fig. 13). Mechanistically, our findings show that depletion of *ARID1A* results in decreased HDAC1 binding, a gain in histone 4 acetylation and coincident BRD4 recruitment at regulatory elements adjacent to genes normally repressed by ER targeted drugs in wild type contexts (Supplementary Fig. 14). This culminates in increased basal proliferation that occurs in a BET-dependent manner. In support of the intrinsic regulation of proliferation by ARID1A, we assessed breast cancer patients with *ARID1A* mutations, when compared to patients with wild type *ARID1A* [27], revealing a poorer clinical outcome in women with *ARID1A*-mutant tumours (Fig. 6d and Supplementary Fig. 14 with details of the statistical test). To explore the link between BET-driven growth in *ARID1A* null contexts and to assess other treatment options for women with *ARID1A* mutations, we established a tumour explant from an *ARID1A* mutant PDX tumour which has a frameshift mutation leading to *ARID1A* loss [30] (Supplementary Fig. 14). Tumour tissue was cultivated *ex vivo* and treated with vehicle or two different BETi for 48hr and we could show significant antiproliferative effects by assessing Ki67 expression, a surrogate marker for proliferation, following treatment (Figs. 6e and 6f), confirming the dependence on BET proteins in *ARID1A* mutant/deleted contexts similar to wild type contexts.

Our study shows that the SWI/SNF complex is recruited to ER cis-regulatory elements prior to active ER binding, via the pioneer factor FOXA1. ARID1A exhibits transcriptional repression by

recruiting HDAC1 and when ARID1A is functionally inactivated, HDAC1 binding is diminished, resulting in a gain in enhancer-specific acetylation, which is subsequently 'read' by BET proteins (Fig. 6g and Supplementary Fig. 15).

Discussion

Our unbiased genetic screening approach has revealed a critical role for the SWI/SNF complex in estrogen receptor-targeted treatment efficacy. Loss of *ARID1A* had profound effects on the gene expression program and growth phenotype, by affecting the chromatin environment. Tumour growth and clinical outcome were influenced by ARID1A status, independent of estrogen-bound ER activity, in support of previous work showed that BAF57 could be recruited to the ER target gene promoter, pS2 (TFF1) in an estrogen independent manner [12]. In contrast, Glucocorticoid Receptor (GR) was shown to recruit the BAF complex to the MMTV chromatin template [8,32], implying that the mode of BAF-chromatin occupancy is nuclear receptor-specific. Our findings suggest that while ARID1A and SWI/SNF components can be recruited to ER cis-regulatory elements by ER antagonistic ligands, in particular Tamoxifen, this complex can associate with these enhancer elements independent of nuclear receptor activation. In this study, we identified that the pioneer factor FOXA1, which demarcates ER regulatory elements [2,33] and binds chromatin independently of hormonal status, is responsible for recruiting the BAF complex to the chromatin. FOXA1 can directly recruit the histone modifying methyltransferase that deposits the histone modification that is the hallmark of enhancer elements [34]and previous work has shown that FOXA1 can open a compacted chromatin template, independently of other proteins [35,36], placing it upstream of all factors that subsequently get recruited to these enhancer elements.

Mutation of ARID1A occurs in ~5% of primary breast cancer, but the frequency increases to ~12% when looking in the metastatic context [20], implying a selection for tumour cells possessing loss-of-function *ARID1A* mutations [20,21]. Our findings suggest that loss of *ARID1A* causes a shift

in the H4 acetylation status, a result of decreased HDAC1 binding, which consequently results in BRD4 recruitment and BET-dependent growth (Fig. 6g). Since *ARID1A* (and other components of the BAF complex) is commonly mutated in many cancer types, a role for this complex in regulating general proliferation status may involve co-opting the key cell type-specific *cis*-regulatory elements. Recent studies highlighted the possibility of exploiting a synthetic lethality-based treatment strategy in *ARID1A*-mutant ovarian cancers, using inhibitors of BET proteins [37,38]. BET inhibitors are proven to be effective in ER-dependent breast cancer cells [25] and our current work implies an increased dependency on epigenetic readers that drive cell division when the activity of the BAF complex is compromised. Given the frequency of BAF mutations in breast cancer, particularly drug resistant contexts, our findings would suggest exploring the potential of epigenetic inhibitors that target the BET proteins.

Acknowledgments

The authors would like to thank Genomics, Proteomics, Bioinformatics, Preclinical genome editing, Histopathology and the Flow cytometry core facilities at Cancer Research UK Cambridge Institute. In particular, we thank C. D'Santos from Proteomics, A. Smith from Preclinical genome editing and M. Clayton, G. Cronshaw, R. Ellis and all the staff in CRUK CI BRU. We acknowledge the support of the University of Cambridge and Cancer Research UK. pSpCas9(BB)-2A-GFP (PX458) used for making CRISPR knockout clones was a kind gift from F. Zhang, Broad Institute, Boston (Addgene # 48138) and the BRD4 antibody from Prof. Cheng Ming Chiang, UT Southwestern Medical Center, Texas. We acknowledge the suggestions from Z. Najafova for the mice experiments and A. Nicholson, M. Sen, X. Wang and S. A. Johnsen for reagents and antibodies. We thank R. Rony for his help in graphic design and models. J.S. Carroll is supported by an ERC Consolidator award (Project number 646876), a Susan G Komen leadership grant and CRUK core funding. S. Nagarajan is supported by Komen grant.

S.V. Rao is funded by Innovation postdoc granted by the Norwegian Research Council. R. Siersbæk is funded by the Novo Nordisk Foundation (NNF15OC0014136).

Author contributions

S.N contributed on conceptualization, methodology, experimental work, formal analysis and figure assembly, writing of text, reviewing and advising on the manuscript. S.V.R contributed on experimental work, analysis and figure assembly of mice experiments and advising on the manuscript. J.S contributed on experimental work, analysis and figure assembly of mice experiments and advising on the manuscript. D.C contributed on experimental work and advising on the manuscript. S.D contributed on methodology and advising on the manuscript. E.K.P contributed on experimental work, formal analysis and figure assembly and advising on the manuscript. J-E.G.P contributed on experimental work and advising on the manuscript. D-L.C contributed on statistical analyses, figure assembly, reviewing and advising on the manuscript. S.K contributed on methodology and advising on the manuscript. K.K contributed on bioinformatic analyses, figure assembly, and advising on the manuscript. C.S.R.C contributed on bioinformatic analyses, figure assembly, and advising on the manuscript. S-E.G contributed on methodology and advising on the manuscript. E.A.G contributed on experimental work and advising on the manuscript. C.B contributed on histopathological analyses and advising on the manuscript. N.G performed ARID1A immunohistochemistry and advising on the manuscript. R.N performed ARID1A immunohistochemistry and advising on the manuscript. A.B contributed on methodology regarding PDX material and advising on the manuscript. C.C contributed on methodology regarding PDX material, funding acquisition and advising on the manuscript. A.R methodology and advising on the manuscript. R.S contributed on methodology, supervision and advising on the manuscript. K.Y contributed on methodology, funding acquisition and advising on the manuscript. I.C contributed on methodology, bioinformatic analyses and figure assembly, writing of text and advising on the manuscript. J.S.C contributed

on conceptualization, supervision, funding acquisition, writing of text, reviewing and advising on the manuscript.

Declaration of Interests

Jason S. Carroll is the founder and CSO of Azeria Therapeutics.

References

1. Ali, S. & Coombes, R.C. Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer* **2**, 101-12 (2002).
2. Carroll, J.S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33-43 (2005).
3. Eeckhoute, J. *et al.* Positive cross-regulatory loop ties GATA-3 to Estrogen Receptor alpha expression in breast cancer. *Cancer Res.* **67**, 6477-83 (2007).
4. Musgrove, E.A. & Sutherland, R.L. Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* **9**, 631-43 (2009).
5. Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843-52 (2000).
6. Malovannaya, A. *et al.* Analysis of the human endogenous coregulator complexome. *Cell* **145**, 787-99 (2011).
7. Mohammed, H. *et al.* Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep* **3**, 342-9 (2013).
8. Fletcher, T.M. *et al.* ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. *Mol Cell Biol* **22**, 3255-63 (2002).
9. John, S. *et al.* Interaction of the glucocorticoid receptor with the chromatin landscape. *Mol Cell* **29**, 611-24 (2008).
10. Michel, B.C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol* **20**, 1410-1420 (2018).
11. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* (2018).
12. Belandia, B., Orford, R.L., Hurst, H.C. & Parker, M.G. Targeting of SWI/SNF chromatin remodelling complexes to estrogen-responsive genes. *Embo J* **21**, 4094-103 (2002).
13. Garcia-Pedrero, J.M., Kiskinis, E., Parker, M.G. & Belandia, B. The SWI/SNF chromatin remodeling subunit BAF57 is a critical regulator of estrogen receptor function in breast cancer cells. *J Biol Chem* **281**, 22656-64 (2006).
14. Jeong, K.W., Lee, Y.H. & Stallcup, M.R. Recruitment of the SWI/SNF chromatin remodeling complex to steroid hormone-regulated promoters by nuclear receptor coactivator flightless-I. *J Biol Chem* **284**, 29298-309 (2009).
15. DiRenzo, J. *et al.* BRG-1 is recruited to estrogen-responsive promoters and cooperates with factors involved in histone acetylation. *Mol Cell Biol* **20**, 7541-9 (2000).
16. Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
17. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* **45**, 592-601 (2013).

18. Wei, Z. *et al.* Vitamin D Switches BAF Complexes to Protect beta Cells. *Cell* **173**, 1135-1149 e15 (2018).
19. Cho, H.D. *et al.* Loss of Tumor Suppressor ARID1A Protein Expression Correlates with Poor Prognosis in Patients with Primary Breast Cancer. *J Breast Cancer* **18**, 339-46 (2015).
20. Yates, L.R. *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32**, 169-184 e7 (2017).
21. St Pierre, R. & Kadoch, C. Mammalian SWI/SNF complexes in cancer: emerging therapeutic opportunities. *Curr Opin Genet Dev* **42**, 56-67 (2017).
22. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **7**, 11479 (2016).
23. Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep* **17**, 1193-1205 (2016).
24. Shi, J. & Vakoc, C.R. The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol Cell* **54**, 728-36 (2014).
25. Nagarajan, S. *et al.* Bromodomain protein BRD4 is required for estrogen receptor-dependent enhancer activation and gene transcription. *Cell Rep* **8**, 460-9 (2014).
26. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
27. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-52 (2012).
28. Glont, S.E. *et al.* Identification of ChIP-seq and RIME grade antibodies for Estrogen Receptor alpha. *PLoS One* **14**, e0215340 (2019).
29. Papachristou, E.K. *et al.* A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nat Commun* **9**, 2311 (2018).
30. Bruna, A. *et al.* A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell* **167**, 260-274 e22 (2016).
31. Filippakopoulos, P. *et al.* Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214-31 (2012).
32. Johnson, T.A., Elbi, C., Parekh, B.S., Hager, G.L. & John, S. Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. *Mol Biol Cell* **19**, 3308-22 (2008).
33. Augello, M.A., Hickey, T.E. & Knudsen, K.E. FOXA1: master of steroid receptor function in cancer. *EMBO J* **30**, 3885-94 (2011).
34. Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S. & Carroll, J.S. FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell Rep* **17**, 2715-2723 (2016).
35. Cirillo, L.A. & Zaret, K.S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4**, 961-9 (1999).
36. Cirillo, L.A. & Zaret, K.S. Specific interactions of the wing domains of FOXA1 transcription factor with DNA. *J Mol Biol* **366**, 720-4 (2007).
37. Berns, K. *et al.* ARID1A mutation sensitizes most ovarian clear cell carcinomas to BET inhibitors. *Oncogene* **37**, 4611-4625 (2018).
38. Caumanns, J.J., Wisman, G.B.A., Berns, K., van der Zee, A.G.J. & de Jong, S. ARID1A mutant ovarian clear cell carcinoma: A clear target for synthetic lethal strategies. *Biochim Biophys Acta Rev Cancer* **1870**, 176-184 (2018).

Figure legends

Fig. 1. CRISPR screens reveal *ARID1A* and BAF components as essential genes for treatment response. Log2 fold of gRNA counts change as a function of time per gene (red lines) and on average (black line) based on a sample of n=3 for three categories of genes: the ones showing a rapid growth depletion (**a**), the ones showing a longer-term growth depletion (**b**) and the ones showing increased proliferation (**c**). For each category, example genes are shown in red and *ARID1A* is shown in blue. **d**. Heatmap representing log2 fold change of significant genes (n=1915) in non-treated conditions (day 3 to day 20 of infection comparing to uninfected gRNA pool). Rows were ordered according to hierarchical clustering. **e**. Heatmap representing log2 fold change of genes after 26 days of treatment with Fulvestrant (Fulv, initiated with 300 nM and reduced to 100 nM gradually), 100 nM 4-hydroxytamoxifen (Tamox) or BETi (JQ1 – 1μM reduced to 250 nM) comparing to DMSO treatment (DMSO control after day 9 of infection). Rows were ordered according to hierarchical clustering. **f**. *ARID1A* and other BAF components were enriched, but in different directions depending on the specific drug treatment. The values show changes in gRNA levels for these genes, using a log2 fold change relative to DMSO control. **g**. Frequency of single gRNAs in log 2 scale against BAF complex subunits *ARID1A, ARID1B, SMARCB1 and SS18*, comparing 4-hydroxytamoxifen or JQ1 with non-treated conditions. **h**. Example of ARID1A ChIP-seq binding overlap with ER and FOXA1 from MCF7 cells grown in media containing 10% fetal bovine serum containing estrogen, from three independent biological ChIP-seq samples per group. **i**. Global overlap between ARID1A, ER and FOXA1 ChIP-seq data from MCF7 cells grown in media containing 10% fetal bovine serum containing estrogen (n=3 independent biological ChIP-seq samples per group). **j**. Heatmaps representing ARID1A binding in hormone-deprived cells treated with vehicle or 10 nM estrogen (n = 3) on the constant sites (n=24,754 sites) defined by DiffBind without any significant change with estrogen treatment and the DiffBind-defined significant sites (n=3,023) which show reduced ARID1A binding during estrogen treatment. Also shown are the relative ER and FOXA1 binding intensities at these regions.

Fig. 2. *ARID1A* knock-out clonal cells show loss of response to ER antagonists, but responsiveness to BET inhibitors. a. *ARID1A* was knocked-out of MCF7 cells using CRISPR deletion. Western blots of ARID1A or ER confirm effective gene deletion in clones 11 and 14, with no change in total ER levels. This figure shows the data of one representative experiment (Source Data Fig. 2) out of the three independent experiments. **b**. Percentage confluence as a function treatment time, in an *in-vitro* proliferation assay using Incucyte conducted in asynchronous MCF7 cells treated with vehicle or 1 μM 4-hydroxytamoxifen. This figure shows the data of one representative experiment out of the four independent experiments. Each experiment considered n=3 replicated per group. Mean ± Standard error of the mean is shown in the graph. **c**. Xenograft tumour volume of MCF-7 (n=13 animals), *ARID1A* K.O clone 14 (n=8 animals), *ARID1A* K.O clone 11 (n=12 animals) as a function of time since day of enrolment. The dots and arrows respectively show the average tumour volume and corresponding 95% confidence intervals of mice at risk. Tumour size of animals at different time-points were fitted by means of a linear mixed model on the cubic root scale, with time and group as fixed effect and random intercepts and slopes for mice (Full details are provided in Supplementary Note). The colored curves and shaded areas correspond to the fitted growth curves for each group and 95% confidence intervals, and the p-values to the mixed model difference in growth rate tests. Test statistics in Fig. S5d. p-values were calculated by two-sided Wald test. **d**. RNA-seq was conducted on the *ARID1A* knockout cells treated with Vehicle, 10 nM Fulvestrant, 100 nM 4-hydroxytamoxifen or 250 μM JQ1 (n=4 independent biological samples). As controls, both parental MCF7 cells and three wild type clonal lines were included. The plot shows fold change of Fulvestrant-regulated genes (n=1094) (ordered by means of a hierarchical clustering) in wild type cells. Highlighted gene cluster (with a star) shows the maintained downregulated effect of JQ1 regardless of *ARID1A* status, but upregulation with Vehicle and 4-hydroxytamoxifen upon ARID1A loss. **e**. Survival rate as a function of time-to-event for 2 groups of ER+ cancer patients: patients showing up- (red) (n=104 for Vehicle and 72 for 4-hydroxytamoxifen) and down- (blue)

(n=101 for Vehicle and 61 for 4-hydroxytamoxifen) regulation according to a gene signature defined by ARID1A targeted genes shown to be repressed by vehicle or 4-hydroxytamoxifen. p-values correspond to log-rank tests (two-sided) (estimated test statistics available in Supplementary Fig. 6) respectively comparing the survival distribution of patients with up and down - regulated genes. Total METABRIC cases: 1181.

Fig. 3. The SWI/SNF complex interacts with ER and is recruited to chromatin following drug treatment. a. ER, ARID1A or BRG1 RIME was conducted in asynchronous MCF7 cells. IgG was used as a negative control. ER, FOXA1 and HDAC1 were identified as interactors in the ARID1A and BRG1 pull downs and vice versa. Boxplots shows the enrichment of selected known interactors in the pulldown samples compared to IgG controls. Pull downs were performed in two biological cell culture samples and label free quantification was performed using Minora algorithm. The log2 intensities are normalised so that the median of IgGs is zero. Centre line shows the median. n=2 independent biological cell culture samples. **b**. Five ER+ PR+ primary tumour samples were split for ER or IgG pull downs and the enrichment of known co-factors in the ER compare to IgGs such as HDAC1 and BAF components are shown. Boxplots shows the enrichment of selected known ERα interactors in the ERα RIME samples compared to IgG controls in human breast cancer tissues. The log2 values are normalised so that the median of IgGs is zero. Centre line shows the median. **c**. ER qPLEX-RIME was conducted in asynchronous MCF7 cells treated with 100 nM 4-hydroxytamoxifen in a 4-point time course (n = 6 independent biological samples per group). Specific BAF proteins are highlighted and the enrichment of the BAF components in the ER complex upon 4-hydroxytamoxifen treatment is shown. Centre line shows the median. **d**. ChIP-seq of ARID1A, BRG1 or SNF5 (*SMARCB1*/BAF47) in asynchronous MCF7 cells treated with vehicle (ethanol) or 100 nM 4-hydroxytamoxifen (n = 3 independent biological ChIP-seq samples). The heatmaps

represent the 39,214 ARID1A binding events observed after 4-hydroxytamoxifen treatment. Also included are H3K27Ac, ER and FOXA1 binding signal intensity at these regions.

Fig. 4. FOXA1 promotes binding of ARID1A and BRG1 to a subset of potential enhancer elements. a, b. Hormone-deprived MCF7 cells were transfected with control or *FOXA1* siRNA and ChIP-seq was conducted for ARID1A (**a**) or BRG1 (**b**). n = 3 independent biological ChIP-seq samples. MA plots are shown with the average intensity of binding vs log2 fold change with *FOXA1* siRNA comparing to control siRNA. **c, d**. Heatmaps (**c**) and boxplots (**d**) shown on ARID1A-BRG1 constant (n= 65563 sites) and ARID1A-BRG1 lost sites (n= 9355 sites) defined by DiffBind following *FOXA1* silencing in MCF7 cells. ER and FOXA1 overlap are also shown on (**c**) these sites. n = 3 independent biological cell culture samples. p-values (**d**) were calculated by Welch's t-test, two-sided. For boxplot, centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range). More statistical details are mentioned in Supplementary Table 5a. **e**. Scatterplot showing the association of decreased ARID1A and BRG1 binding following *FOXA1* silencing. PCC – Pearson Correlation coefficient. p-values were calculated by Pearson Correlation test, two-sided. **f-g**. Scatterplot showing the association of ARID1A (**f**) and BRG1 (**g**) binding following *FOXA1* silencing at tamoxifen-induced ARID1A (**f**) and BRG1 (**g**) binding sites from Fig. 3d. PCC – Pearson Correlation coefficient. p-values were calculated by Pearson Correlation test, two-sided. **h**. Boxplots illustrating the effect of si*FOXA1* on ARID1A and BRG1 binding on the ER binding sites (n=2,746 sites) close to ARID1A repressed genes in Vehicle conditions. p-values were calculated by Welch's t-test, two-sided. Window – 400 bp around center of the factor binding. Centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further

than 1.5 × IQR (inter-quartile range). More statistical details are mentioned in Supplementary Table 5b.

Fig. 5. Loss of ARID1A results in decreased BRG1 and HDAC1 recruitment and increased histone H4 acetylation. a, b, c. Quantitative signal from BRG1 (**a**), HDAC1 (**b**) and ER (**c**) ChIP-seq within *ARID1A* knock-out cells (n=3 independent biological cell culture samples per group). ChIP-seq was conducted in the wild type cells or the two *ARID1A* knock-out clones, showing decreased binding of the factors in the absence of ARID1A. Average plots were shown on HDAC1 lost sites in the *ARID1A* knock-out cells. **d, e**. Scatterplots showing the association of decreased BRG1 and HDAC1 binding in *ARID1A* knockout clone 11 (**d**) and clone 14 (**e**) following 100 nM 4-hydroxytamoxifen treatment. n = 3 independent biological cell culture samples. PCC – Pearson Correlation coefficient. P-values were calculated by Pearson Correlation test, two-sided. **f**. ER qPLEX-RIME was conducted in four ER+ PDX tumours, two of which had loss of ARID1A via mutation (MT1/2) and two were wild type (WT1/2) for *ARID1A*. Heatmaps reveals decreased BAF and HDAC1 interactions with ER in *ARID1A* mutant tumours compare to the wild type tumours. **g**. We specifically identified ARID1A repressed genes in proximity to the ER-bound regulatory elements (n=686 sites) that display, according to PCA, more than 75% contribution to the variance in intensity of histone H4 acetylation. The data is shown as boxplots. ARID1A dependent genes acquired gained H4 acetylation, especially H4K8Ac and H4K12Ac at adjacent enhancers, coincident with increased gene expression. P-values were calculated by Welch's t-test, two-sided. Window – 2 kb around center of the binding event. More statistical details are provided in Supplementary Table 5c.

Fig. 6. Loss of ARID1A results in increased BRD4 recruitment and a gain in intrinsic proliferation. a- c. BRD4 ChIP-seq was conducted in wild type or *ARID1A* knock-out cells (n=3 independent ChIP-sew samples per group). **a**. Boxplots were shown on ER bound regions close to ARID1A repressed genes (n=686) that display, according to PCA, more than 75%

contribution to the variance in intensity of H4 acetylation. p-values were calculated by Welch's t-test, two-sided. Window – 400 bp around center of the factor binding. For boxplot, centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range). More statistical details are mentioned in Supplementary Table 5d. **b**. Scatterplot showing the association of BRD4 and HDAC1 binding in the *ARID1A* knockout clone 11 cells. n = 3 independent biological cell culture samples. PCC – Pearson Correlation coefficient. P-values were calculated by Pearson Correlation test, two-sided. **c**. Heatmap shows the gained BRD4 occupancy and decreased HDAC1 binding on the ER-bound regulatory elements (n=2,746 sites) adjacent to ARID1A target genes. **d**. Overall patient survival was assessed based on *ARID1A* mutational status in a cohort of 1,824 breast cancer patients. p-value was calculated by log rank survival test, two-sided (estimated test statistics available in Supplementary Fig. 14). **e**. Ki67 IHC protein levels stained on an *ARID1A* mutant PDX cultivated *ex vivo* and treated with DMSO vehicle (n=10 explant chunks), 250 nM JQ1 (n=9 explant chunks) or 1μM IBET762 (n =10 explant chunks) for 48hr in a single experiment. Median values are shown with p-values calculated using Wilcoxon test, two-sided (Wilcoxon test Statistic W= 17 for both comparisons). **f**. Representative images of Ki67 IHC in BETi ex vivo tumour tissue, with each image representing a region of 100 μm in length. The explant chunks were treated with DMSO vehicle (n=10), 250 nM JQ1 (n=9) or 1μM IBET762 (n =10). **g**. Model of FOXA1-ARID1A-HDAC1-BRD4 axis in *ARID1A* wild type and mutant contexts.

Online Methods

The Life Sciences Reporting Summary for this manuscript is provided in Supplementary information.

Preparation of Cas9-expressing clones

MCF7 cells were transduced with Cas9 lentiviral vector pKLV2-EF1aCas9T2ABsd-W with 8 ug/ml Polybrene in 2% serum containing media without antibiotics. Media was replaced after 24 hrs with 10% serum, grown for two more days and selected with 30 µg/ml Blasticidin for four days. These cells were single cell sorted using (BD FACSAria II) in one 96 well plate, seeded with very high suspension and diluted into two 15 cm dishes and grown in the presence of Blasticidin. After 10 days of growth, single cell clones were hand-picked and seeded and grown in two 96 well plates. After the clones were grown well, 48 clones were selected and assessed for Cas9 cutting efficiency using reporter assay in a 6 well plate. Cas9 clones were transduced separately with pKLV2-U6gRNA5(GFPg0)-PGKBFPGFP-W where the cells can express BFP and GFP (control) after 3 days of infection and pKLV2-U6gRNA5(GFPg5)-PGKBFPGFP-W which has a gRNA for GFP. Highly efficient clones were selected which shows ~95% BFP+ cells in the infected population sorted by BD Influx™ Cell Sorter (Supplementary Fig. 1). The average efficiency was calculated from 4 independent experiments. Cas9 cut efficiency is calculated as follows:

Cas9 efficiency (%) = 100 – (%. of GFP+BFP+cells/(%. of GFP-BFP+cells + %. of GFP+BFP+cells)*100) which means 100 – (%. of uncut cells/Total % of transfected cells*100)

1C3 clone was selected from FACS-sorted plate and showed 93.9% Cas9 cut efficiency. This had been used for initial essentiality screen. 3G1 clone sorted from highly diluted plate showed 94.62% Cas9 cut efficiency which was used for drug resistance screening. Both the clones were compared after infection with gRNA library after 9 days for their reproducibility.

Genome-wide CRISPR screening

Highly efficient Cas9-expressing cells were infected with the human gRNA pooled library version-1 with the vector backbone pKLV2-U6gRNA5(lib)-PGKpuroBFP-W[23]. Cells were seeded two days before in a 15 cm dish to ~30 million cells per replicate. Cells were infected with 30%

transduction efficiency (Supplementary Fig. 1) so that only one gRNA gets integrated into the genome per cell. After 3 days (D3), the 30% transduction efficiency was verified by FACS sorting in Influx for BFP+ cells. 60 million cells were collected for next generation sequencing and antibiotic selection was performed on the remaining cells with 10 µg/ml Puromycin for 4 days. BFP+ cells were at least 95% after 4 days of antibiotic selection which was verified by FACS sorting in Influx. Consequently, 100 million cells were collected on different number of days (Day 7, 9, 12, 15 and 20). Genomic DNA was isolated from 20-50 million cells using Qiagen Blood & Cell Culture DNA Maxi Kit along with RNase treatment.

a. Drug resistance CRISPR screening

After 9 days of infection with gRNA library, cells were treated with 100 nM 4-hydroxy-tamoxifen. Fulvestrant and JQ1 were used at 300 nM and 1 µM, respectively, at the start of the assay and gradually reduced to 100 nM and 250 nM, respectively. All treatments were done for 26 days. DMSO was used as a control.

b. Library preparation for CRISPR screens

90 µg of genomic DNA from CRISPR library-infected cells which represents 10 million MCF7 cells (100X representation of gRNA library) were amplified as 5 ug per reaction (20 times) using primers with lentiviral and Illumina adapter sequences with Q5 Hot Start High-Fidelity 2X Master Mix (New England biolabs). Primers were noted in Supplementary Table 6. 15 ng of the whole-genome plasmid library per reaction was used corresponding to 1.7×10^{10} molecules of the plasmid DNA. PAGE-purified primers (Sigma) were used: PCR was done for 25 cycles as follows: Initial denaturation 98°C, 30sec; denaturation 98°C, 10sec: annealing 61°C, 15sec; extension 72°C, 20sec; final extension 72°C, 2min. The PCR reaction was verified using Agarose gel electrophoresis for the presence of 250 bp PCR product. 5 µl from each reaction was taken, pooled and purified using Qiagen PCR purification kit. Second amplification was

performed on 100 ng of PCR-purified DNA using Illumina dual indices from Takara ThruPLEX DNA-seq 96D Kit R400407 and KAPA HiFi HotStart ReadyMix for 8 cycles as follows: initial denaturation 98°C, 30sec; denaturation 98°C, 10sec; annealing 66°C, 15sec and extension 72°C, 20sec; final extension 72°C, 5min. Final PCR product was purified using Beckman Coulter Agencourt Ampure XP beads with 0.7X ratio. Libraries were checked for size by Agilent Bioanalyser 2100 or Tapestation 4000 and quantified by qPCR using KAPA library quantification kit with ROX Low qPCR Master Mix or Qubit 3.0 Flurometer. These were pooled and sequenced using an Illumina HiSeq 4000 with 50 bp single end reads with 30% Ilumina PhiX Control spike-in version 3. 30 million reads per sample to sequence every gRNA from every independently infected cell.

c. CRISPR screening analyses

Short reads were depleted from low-quality sequences and aligned to human gRNA sequence library (GRC h37) using BLAT v. 34 [39]. Exact-matching reads were counted and treated as a measurement of gRNA abundance. Gene ranking was performed using MAGecK (Full data is available in the Supplementary Table 1 and 2) and log2 fold changes were calculated using DESeq. Heatmaps were generated using median log2 fold changes values from gRNAs specific to a gene. Plasmid library was used as the control for essential gene screening and DMSO for drug resistance screening. Time series clustering was performed using dtwclust R package on genes which showed significant enrichment or depletion. For the heat-map, Dynamic Time Warp algorithm from dtw R package was used for distance measurement followed by hierarchical clustering.

RNA-sequencing

RNA quality was checked using RNA Integrity Number (RIN) from Bioanalyser and 500 ng of RNA was used to prepare libraries using Illumina TruSeq stranded mRNA (HT) library

preparation kit. Library size distribution was assessed using the Agilent Tapestation 4200 system. These were sequenced using HiSeq 4000 50 bp single end sequencing. 1% PhiX version 3 viral genome spike-in was introduced during sequencing. Fastaq single-end reads from multiple lanes were merged to make a single library per replicate. STAR[40], version 2.5.1a, was used to align reads against hg38 reference genome. The read counting was performed using the intrinsic function of STAR. Differential gene expression analysis used the DESeq2 workflow. All p-values were corrected for multiplicity by means of the Benjamini and Hochberg FDR multiplicity correction.

ATAC-sequencing

Omni ATAC-sequencing was performed according to the established protocol from Corces et al [41] . NX# TDE1, Tagment DNA enzyme and buffer from Illumina were used for the transposition reaction. Nextera dual indices were utilized for multiplexing. Sequencing was performed using HiSeq 4000 paired end 150 bp reads.

ChIP-seq and ATAC-seq analyses

Reads were mapped to hg38 genome using bowtie2 2.2.6[42]. Aligned reads with the mapping quality less than 5 were filtered out. The read alignments from three cell culture samples were combined into single library and peaks were called with MACS2 version 2.0.10.20131216[26] using sequences from MCF7 chromatin extracts as a background input control. The peaks yielded with MACS2 q value <= 1e-3 were selected for downstream analysis. Genrich (https://github.com/jsh58/Genrich) was used to verify the ATAC-seq peaks from MACS2. Meme version 4.9.1[43] was used to detect known and discover novel binding motifs amongst tag-enriched sequences.

735 Differential binding analysis (DiffBind) was performed as described previously [44]. For visualizing
736 tag density and signal distribution heatmap the read coverage in a window of +/- 2.5 or 5 kb
737 region flanking the tag midpoint was generated using the bin size of 1/100 of the window length.

738 **Gene signature analysis, KM plots**

739 A set of genes that were evaluated as differentially-expressed in RNA-seq analysis and located
740 in +/- 50kbp vicinity to the differentially-occupied sites evaluated in ChIP-seq analysis was
741 qualified as a potential Gene Signature.

742 METABRIC[27] gene-expression data was accessed via API available at Genomics Data
743 Commons portal (https://gdc.cancer.gov/developers/gdc-application-programming-interface-api
744) ported to MATLAB. Kaplan-Meier plots and log-rank tests were respectively used to display
745 the survival probabilities per group as a function of time and to test if the hazard functions of the
746 groups of interest are different. Groups of clinical cases (n>=20) of BC ER+ cohorts were
747 stratified by expression of group of genes established at a threshold corresponding to most
748 significant difference in survival.

749 **RIME and qPLEX-RIME**

750 **a. RIME on cell lines**

751 Cells were double cross-linked with 2 mM DSG and 1% Formaldehyde as described in ChIP-
752 sequencing in the Supplementary Note. The protocol was followed as in ChIP-seq with following
753 modifications: beads were washed 10 times with RIPA and twice with 100mM ice-cold
754 ammonium hydrogen carbonate. Antibodies used: ARID1A (HPA005456), BRG1 (ab215998),
755 ERα (ab3575 and Merck Millipore 06-935 antibody mix) and negative control IgG (ab171870).

756 **b. qPLEX-RIME on patient-derived xenografts**

Frozen clinical tissues were cryosectioned at 30 micron sections and ~90 sections were double crosslinked with 2mM DSG for 25 mins and 1% formaldehyde in the same solution of DSG for 20 mins. Crosslinking was quenched with 0.25M Glycine. The pull down was performed with the ER antibody mix as mentioned in ChIP-seq and qPLEX-RIME sections.

c. Proteomic sample preparation, LC-MS analysis and Data processing

For sample preparation, trypsin at final concentration 15ng/ul (Pierce) was added to the beads followed by overnight incubation at 37°C. A second digestion step was performed the next day for 4h and peptides were cleaned with the Ultra-Micro C18 Spin Columns (Harvard Apparatus) according to manufacturer's instructions. For the qPLEX-RIME experiment, samples were dried and labelled with the TMT-10plex reagents (Thermo Fisher) followed by fractionation using Reversed-Phase spin columns at high pH (Pierce #84868). For the qPLEX-RIME, peptide fractions were analysed on a Dionex Ultimate 3000 UHPLC system coupled with the nano-ESI Fusion Lumos (Thermo Scientific) mass spectrometer.. The full MS scans were performed in the Orbitrap in the range of 380-1500 m/z at 120K resolution. The MS2 scans were performed in the ion trap with CID collision energy 35%. Peptides were isolated in the quadrupole with isolation window 0.7Th. The top 10 most intense fragments were selected for Synchronous Precursor Selection (SPS) HCD-MS3 analysis with MS2 isolation window 2.0Th. The HCD collision energy was set at 65% and the detection was performed with Orbitrap resolution 50K. For RIME experiments, peptides were analysed on a Dionex Ultimate 3000 UHPLC system coupled with the Q-Exactive HF (Thermo Scientific) or the Q-Exactive mass spectrometers. The full MS scans were acquired in the Orbitrap within the range of 400-1600m/z at 60K or 70K resolution respectively. For MS2, the top 10 most intense precursor ions were selected with a 2.0Th window followed by HCD fragmentation with collision energy 28%. The collected CID and HCD tandem mass spectra were processed with the SequestHT search engine in Proteome Discoverer 2.1 and Proteome Discoverer 2.2 respectively. The SequestHT included the

following parameters: Precursor Mass Tolerance 20ppm, Fragment Mass Tolerance 0.5Da for CID and 0.02Da for HCD, Dynamic Modifications were Oxidation of M (+15.995Da), Deamidation of N/Q (+0.984Da) and Static Modifications were TMT6plex at any N-Terminus/K (+229.163Da) for the qPLEX-RIME experiment only. The consensus workflow included calculation of TMT signal-to-noise and the confidence level for peptide identifications was estimated with the Percolator node with decoy database search. The peptide intensities for the qPLEX-RIME experiment were normalized and aggregated (by summing) to protein intensities. The differential protein expression was performed using limma[45] implemented in the qPLEXanalyzer tool (10.18129/B9.bioc.qPLEXanalyzer). The Minora Feature Detector node implemented on Proteome Discoverer 2.2 was used for label-free quantification at Maximum ΔRT of Isotope Pattern Multiplets 0.2 min and minimum number of isotopes two peaks. The consensus workflow included Feature Mapper and Precursor Ions Quantifier for Precursor Abundance quantification based on intensity. Complete data is available in Supplementary Table 3 and 4.

Sample size calculation for *in vivo* MCF7 xenografts

The sample size of the study was defined so that, based on effect sizes defined on prior data and on nuisance parameters deduced from data of Mohammed *et al* [46], a global power of 0.8 would be achieve when testing a chosen set of differences in means of tumor volumes at the global 5% level for different time points by means of Welsh's tests.

In vivo xenografts

All mouse experiments were carried out in Biological Resource Unit at CRUK Cambridge Institute. The experiments were in accordance with the UK Animals (Scientific Procedures) Act 1986, with approval from the CRUK Cambridge Institute Animal Ethical Review and Welfare Body. Age matched (8 weeks) NOD/SCID/IL2Rg−/− (NSG) female mice were purchased from

Charles River. The animals were verified to be pathogen free and in excellent health. Subcutaneous xenografts of MCF-7 cells/ARID1A clones were conceived by implanting cells 10^5 in 50% growth media and 50% matrigel (BD Biosciences), in the right flank of 8 weeks old female NSG mice. The mice were also implanted subcutaneously with 90 day-slow release 17b-oestradiol (0.72 mg per pellet) hormone pellets (Innovative Research of America) into the left flank. After 4 weeks for the efficacy cohort, the tumors were randomized and enrolled to the study when the average tumor volume was 100-150 mm^3 size. 8.8 mg/ml of Tamoxifen (Tocris Bioscience, 6342) was made in sterile filtered corn oil (Sigma, C8267). The mice were dosed at 20mg/kg, I.P, 6 days a week with Tamoxifen. Tumor sizes were monitored twice a week with Vernier caliper measurement.

As tumor volumes show linear growths on the cubic root scale, we used linear mixed models to compare the average tumor growth of the different groups as a function of time from enrolment on that scale. Linear mixed models allow to take both the within-mouse and time-dependence into account by means of random effects and auto-regressive parameters respectively. We considered here a random intercept and slope model with time since enrolment, groups and an interaction between time since enrolment and groups as fixed effects, and an autocorrelation structure of order 1 for the error term. Model checks suggested a good fit of the model to the data. Sensitivity analyses considering alternative modelling (like models including quadratic terms, other kind of time-dependence or other transformations of the tumor volumes) lead to similar conclusions. We used the program R (version 3.5.1) and the package nlme (version 3.1-137) to fit linear mixed models. Mean values drop for the clones especially at day 18 and 25 as the tumor volume exceeded the 1500 mm^3 limit and were removed from the mice.

Test statistics were shown in Supplementary Fig. S5. In the table S5C, fixed effect estimates of the random intercept and slope model used to fit the growth curves in Figure 2C. The model intercept corresponds to the tumour size of the MCF7 WT group at day of enrolment on the

cubic root scale. The coefficient related to the variable Days correspond to the daily increase in tumour size for the reference group (MCF7) on the cubic root scale. The two last parameters correspond to shift in daily growth of the KO 11 and KO 14 groups compared to MCF7.

Explant culture

The ARID1A mutant Patient-Derived Xenograft AB555B was grown in dental sponges as previously described [47,48]. Spongostan gelatine dental sponges were pre-soaked in explant culture media with or without inhibitors (250 nM JQ1 and 1 µM IBET762) and warmed in a 37°C incubator. One sponge per well was placed in a sterile 24-well tissue culture plate, along with 500 µl explant culture media RPMI 1640 (phenol red-free, L-glutamine-free) (Gibco, 32404-014) with 10% heat inactivated fetal bovine serum (Gibco A3840401), 2mM L-glutamine (Sigma G7513), 10 µg/ml Sigma hydrocortisone (Sigma H0888), 10 µg/ml human recombinant insulin (Sigma I9278)), 100 U penicillin, 100 µg streptomycin, 250 ng amphotericin B /ml (from 1x Sigma anti-biotic, anti-mycotic solution; #A5955). PDX material was cut into 9-12 smaller pieces and each piece was analysed as a replicate. Samples on the sponges were incubated with media with inhibitors for 2 days at 37°C with 5% CO_2. These were collected from sponges and fixed in 10% neutral buffered formalin overnight at room temperature. Tissues were processed and embedded in paraffin for histological assessment. Slides were scanned on an Aperio AT2 (Leica) at 20X magnification (resolution 0.5um per pixel) and analysed using HALO software (Indica labs), with the multiplex IHC v2.1.1 module.

Statistical analyses

Two-sided tests were used for all the statistical analyses. Bar graphs were shown with average values and the box plots with median values. Standard deviation was used to denote the error bars in the bar graphs with average values except the proliferation data from Incucyte assays where standard error of the mean was used. For boxplots, centre line shows the median values

with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range). More details about the boxplots on ChIP-seq data are mentioned in Supplementary Table 5.

Generation of Genome Edited ARID1A Knock Out clones

CRISPR guides (sgRNA) were designed against Exon 2 of ARID1A (NM_006015). Oligos (Sigma Aldrich) were cloned into pSpCas9(BB)-2A-GFP (PX458, Addgene # 48138) as previously described[49]. Guide cutting efficiency was determined in MCF7 and HEK293T cells using the T7 assay (New England Biolabs, following manufacturer's instructions). To generate independent, non-sister clonal cell lines, MCF7 cells were transiently transfected (Lipofectamine 3000, Thermo Fisher Scientific) with PX458-empty (control), PX458-sgARID1A_2.1 and PX458-sgARID1A2.2, and single cell cloned in 5X 96 well plates per gRNA 96h post-transfection by FACS (BD FACSAria II). gDNA was extracted from each clone (Extracta DNA Prep, VWR, cat#95091-025) and Exon 2 of ARID1A was amplified by PCR (FastStart HF System (Sigma Aldrich, cat#3553361001), primers were noted in Supplementary Table 6 (universal Fluidigm tag in lower case). Amplicons were diluted 1:150 and re-amplified with Fluidigm barcoding primers (incorporating a unique sample barcode and Illumina P5 and P7 adapter sequences), pooled and subjected to sequencing (Illumina MiSeq platform). The AmpliconSeq analysis pipeline was used for data processing and variant calling. Briefly, reads were aligned against the reference genome (GRCh38) using BWA-MEM [50] and variants were called using two methods (VarDict[51] and GATK HaplotypeCaller (https://doi.org/10.1101/201178)). Consensus variants and their effects on CRISPR clones were then calculated. All clones used in this paper were STR genotyped and confirmed as free from mycoplasma.

Assessment of off-target CRISPR effects

The top three predicted off targets (ACGGCTCCCTGTCCCCGCAG at chr1:205061276-205061299; AGAGGCCCCAGACCCCGCAG at chr7:1547994-1548017; CCGGCTCCCAGGCCCCGCAG at chr5:10555551-10555574) defined by Desktop Genetics with score 88 out of 100[49] were verified for their absence of editing with Sanger sequencing by amplifying the regions with primers against 3 loci from the final knock-out and empty vector control clones (11, 14, 216, 219, 221). Primers are noted in Supplementary Table 6.

Data Availability Statement

All CRISPR, ChIP-seq and RNA-seq data has been deposited at GEO and can be accessed at GSE123286. ATAC-seq data can be accessed at GSE134270. All proteomic data has been deposited at PRIDE and can be accessed at PXD011810.

Methods-only references

39. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64 (2002).
40. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
41. Corces, M.R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
42. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
43. Bailey, T.L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).
44. Stark, R. & Brown, G.D. DiffBind: differential binding analysis of ChIP-Seq peak data. *Bioconductor* **http://www.http://http://bioconductor.org/packages/release/bioc/html/DiffBind.html** .
45. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
46. Mohammed, H. *et al.* Progesterone receptor modulates ERalpha action in breast cancer. *Nature* **523**, 313-7 (2015).
47. Centenera, M.M., Raj, G.V., Knudsen, K.E., Tilley, W.D. & Butler, L.M. Ex vivo culture of human prostate tissue and drug development. *Nat Rev Urol* **10**, 483-7 (2013).
48. Centenera, M.M. *et al.* A patient-derived explant (PDE) model of hormone-dependent cancer. *Mol Oncol* **12**, 1608-1622 (2018).
49. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).

51. Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* **44**, e108 (2016).

Figure 1

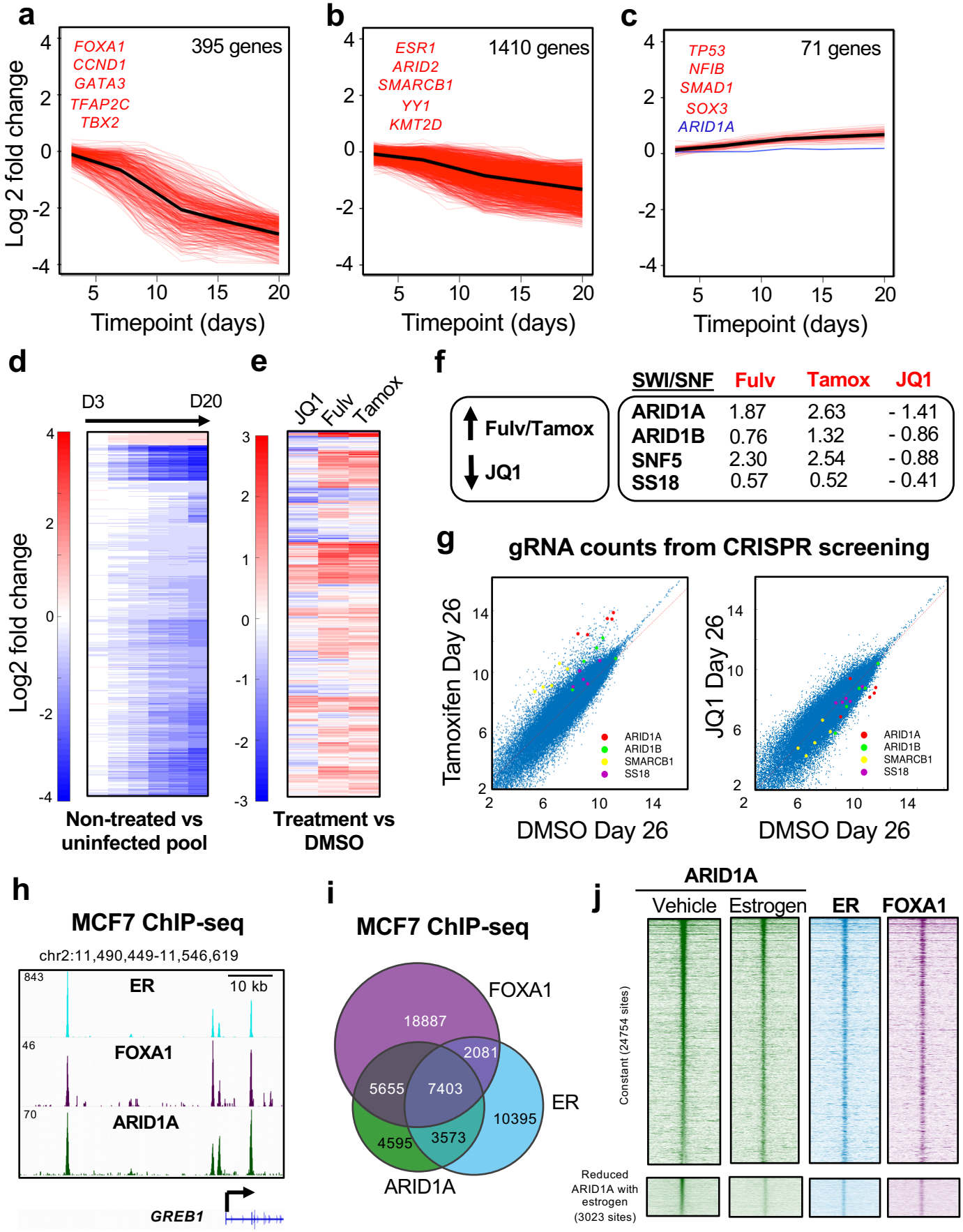

a
FOXA1
CCND1
GATA3
TFAP2C
TBX2
395 genes
Log 2 fold change
Timepoint (days)
b
ESR1
ARID2
SMARCB1
YY1
KMT2D
1410 genes
Timepoint (days)
c
TP53
NFIB
SMAD1
SOX3
ARID1A
71 genes
Timepoint (days)
d
D3
D20
Log2 fold change
Non-treated vs uninfected pool
e
JQ1
Fulv
Tamox
Treatment vs DMSO
f
Fulv/Tamox
JQ1
SWI/SNF Fulv Tamox JQ1
ARID1A 1.87 2.63 - 1.41
ARID1B 0.76 1.32 - 0.86
SNF5 2.30 2.54 - 0.88
SS18 0.57 0.52 - 0.41
g
gRNA counts from CRISPR screening
Tamoxifen Day 26
DMSO Day 26
JQ1 Day 26
DMSO Day 26
ARID1A
ARID1B
SMARCB1
SS18
h
MCF7 ChIP-seq
chr2:11,490,449-11,546,619
ER
FOXA1
ARID1A
10 kb
GREB1
i
MCF7 ChIP-seq
FOXA1
ER
ARID1A
18887
2081
5655
7403
10395
4595
3573
j
ARID1A
Vehicle
Estrogen
ER
FOXA1
Constant (24754 sites)
Reduced ARID1A with estrogen (3023 sites)

Figure 2

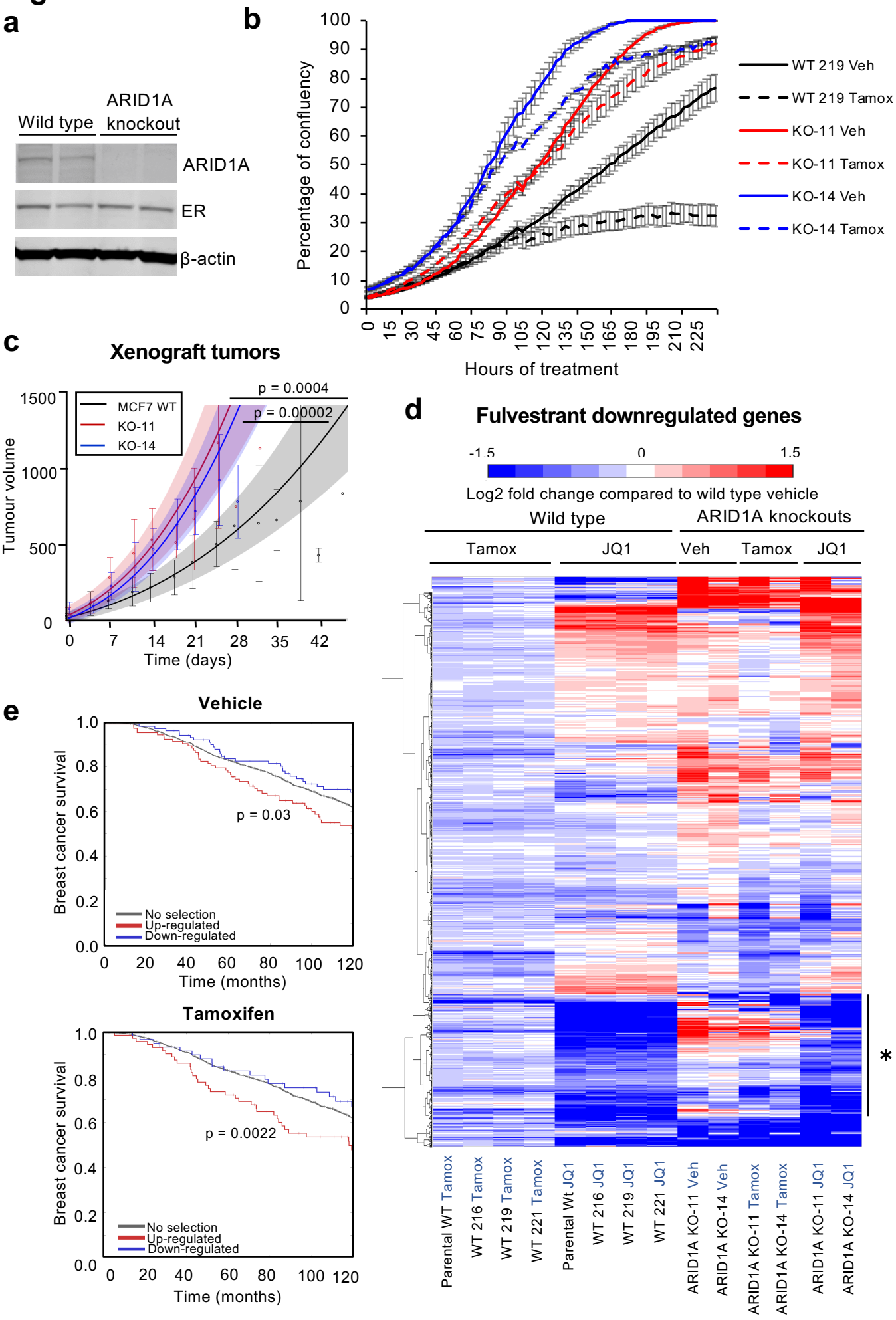
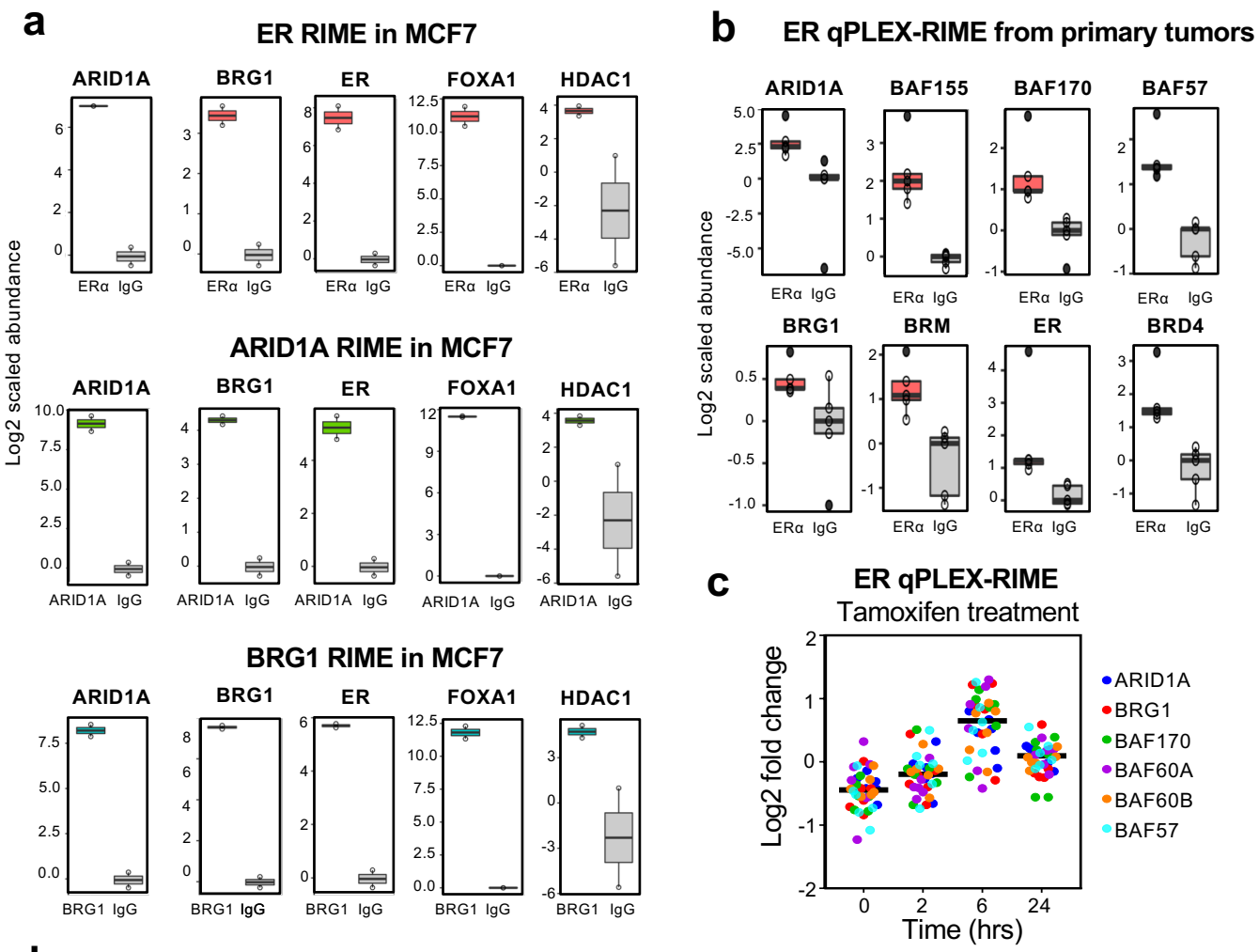

a
Wild type
ARID1A knockout
ARID1A
ER
β-actin
b
Percentage of confluency
Hours of treatment
WT 219 Veh
WT 219 Tamox
KO-11 Veh
KO-11 Tamox
KO-14 Veh
KO-14 Tamox
c
Xenograft tumors
Tumour volume
Time (days)
MCF7 WT
KO-11
KO-14
p = 0.0004
p = 0.00002
d
Fulvestrant downregulated genes
-1.5
0
1.5
Log2 fold change compared to wild type vehicle
Wild type
ARID1A knockouts
Tamox
JQ1
Veh
Tamox
JQ1
Parental WT Tamox
WT 216 Tamox
WT 219 Tamox
WT 221 Tamox
Parental Wt JQ1
WT 216 JQ1
WT 219 JQ1
WT 221 JQ1
ARID1A KO-11 Veh
ARID1A KO-14 Veh
ARID1A KO-11 Tamox
ARID1A KO-14 Tamox
ARID1A KO-11 JQ1
ARID1A KO-14 JQ1
*
e
Vehicle
Breast cancer survival
Time (months)
p = 0.03
No selection
Up-regulated
Down-regulated
Tamoxifen
p = 0.0022

Figure 3


a
ER RIME in MCF7
ARID1A
BRG1
ER
FOXA1
HDAC1
ERα IgG
ARID1A RIME in MCF7
ARID1A IgG
BRG1 RIME in MCF7
BRG1 IgG
Log2 scaled abundance
b
ER qPLEX-RIME from primary tumors
ARID1A
BAF155
BAF170
BAF57
BRG1
BRM
ER
BRD4
ERα IgG
Log2 scaled abundance
c
ER qPLEX-RIME
Tamoxifen treatment
Log2 fold change
Time (hrs)
ARID1A
BRG1
BAF170
BAF60A
BAF60B
BAF57

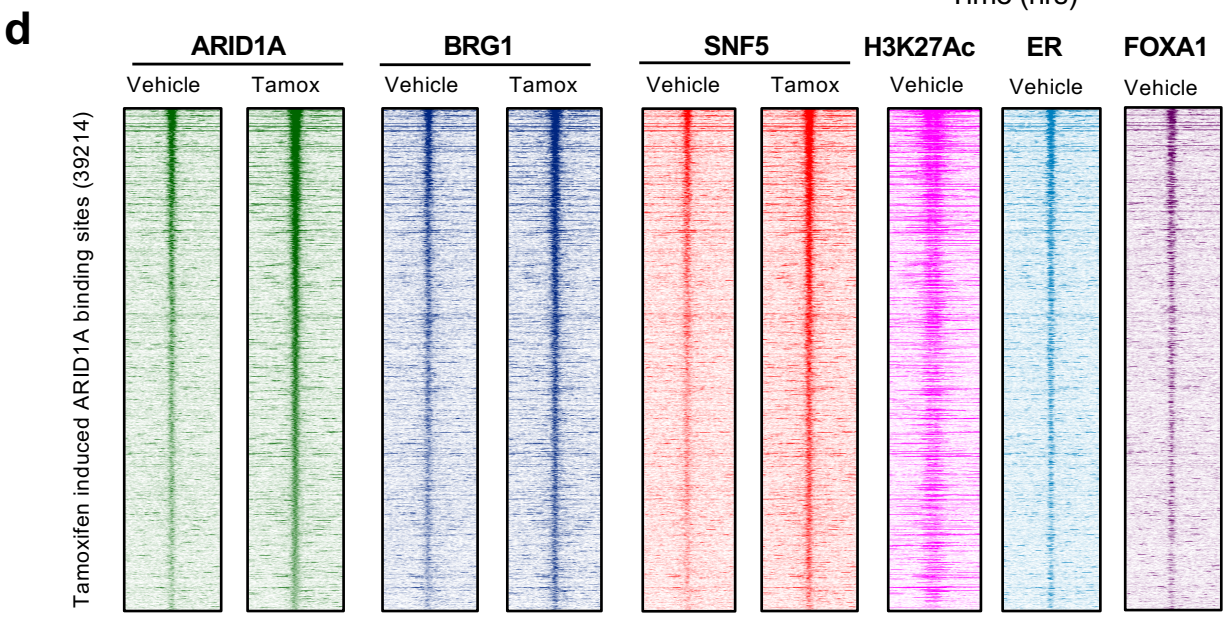

d
ARID1A
BRG1
SNF5
H3K27Ac
ER
FOXA1
Vehicle
Tamox
Tamoxifen induced ARID1A binding sites (39214)

Figure 4

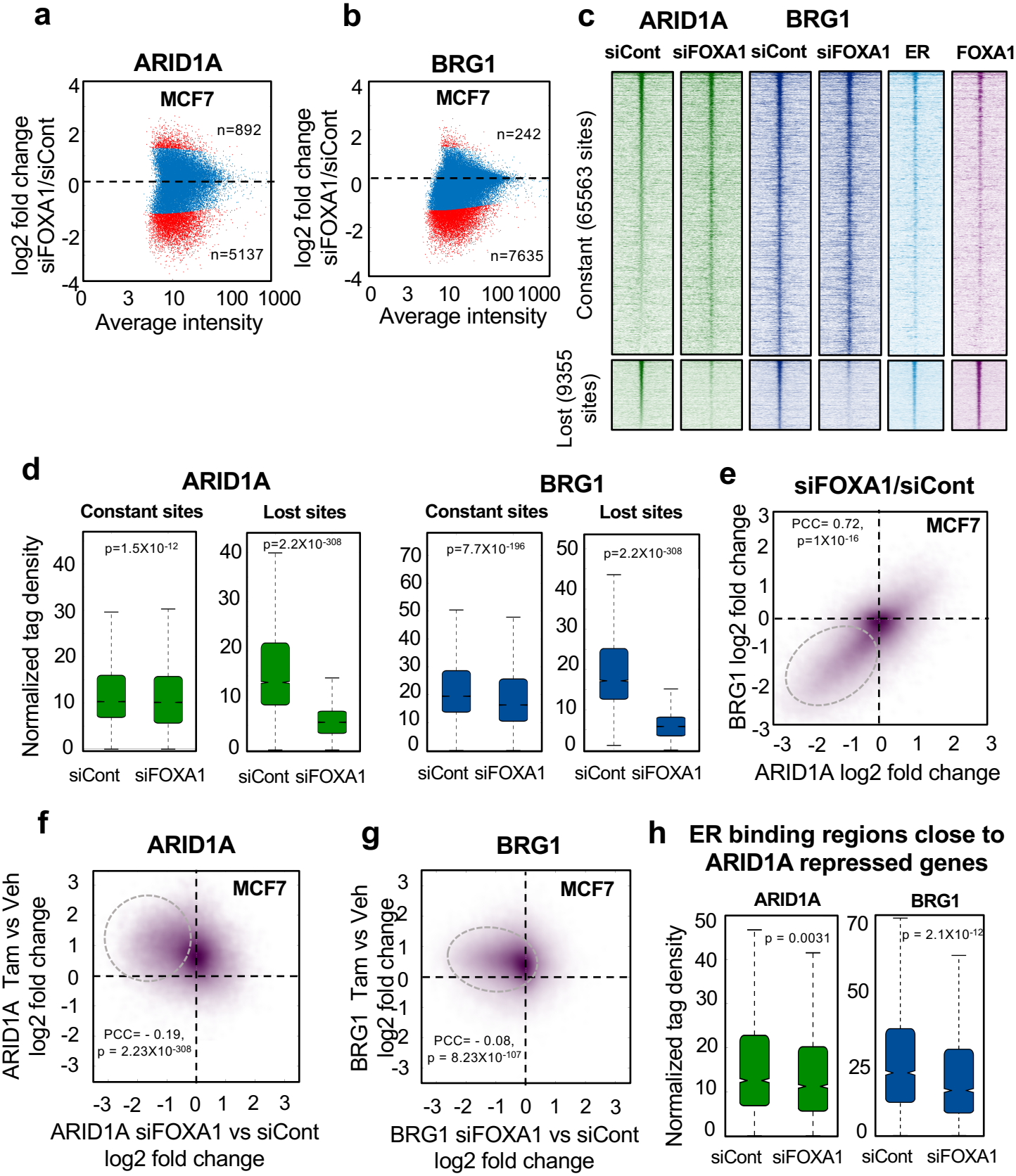
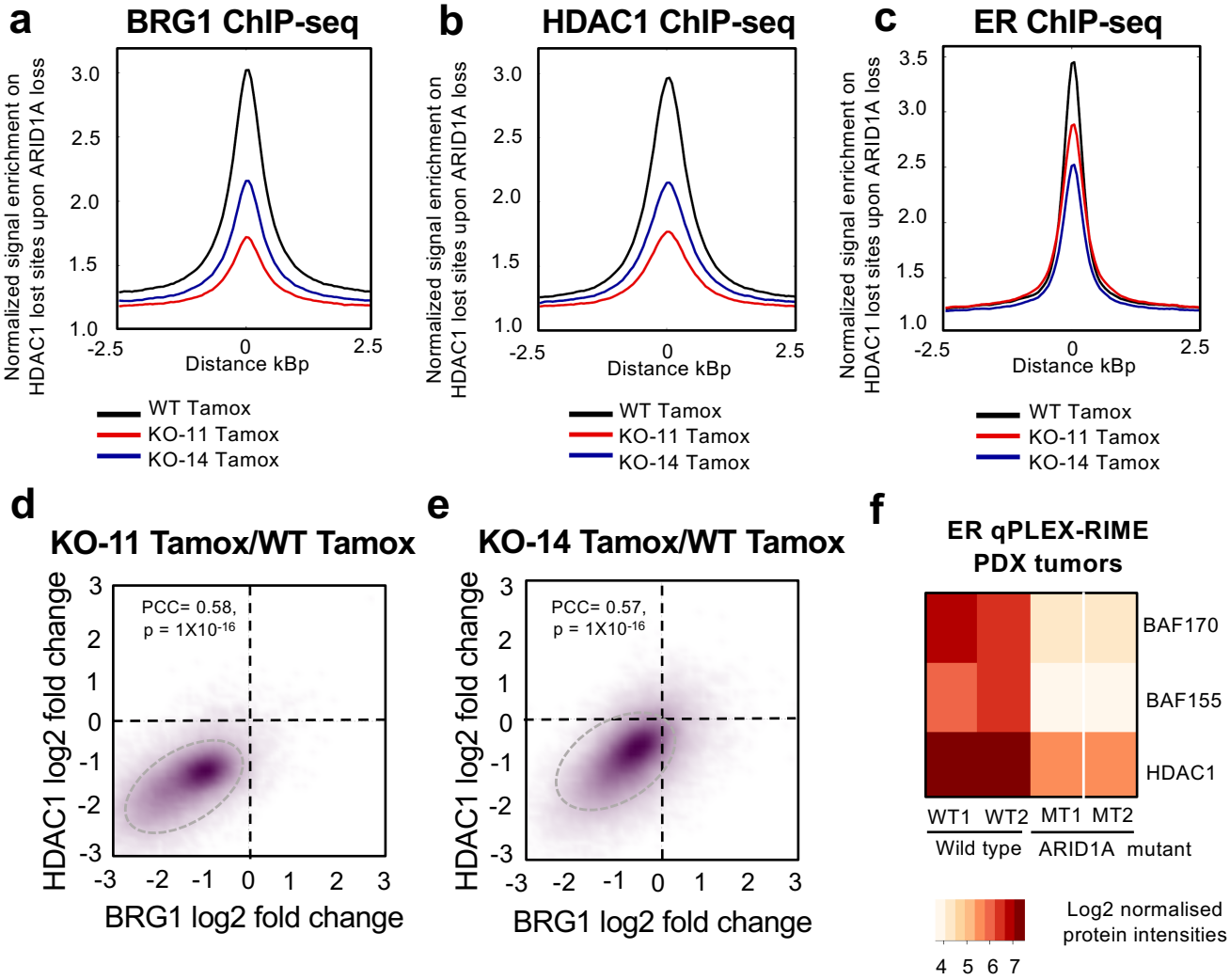

a
ARID1A
MCF7
n=892
n=5137
log2 fold change siFOXA1/siCont
Average intensity
b
BRG1
MCF7
n=242
n=7635
log2 fold change siFOXA1/siCont
Average intensity
c
ARID1A
BRG1
siCont siFOXA1 siCont siFOXA1 ER FOXA1
Constant (65563 sites)
Lost (9355 sites)
d
ARID1A
BRG1
Constant sites
Lost sites
Normalized tag density
p=1.5X10-12
p=2.2X10-308
p=7.7X10-196
p=2.2X10-308
siCont siFOXA1
e
siFOXA1/siCont
MCF7
PCC= 0.72,
p=1X10-16
BRG1 log2 fold change
ARID1A log2 fold change
f
ARID1A
MCF7
PCC= - 0.19,
p = 2.23X10-308
ARID1A Tam vs Veh log2 fold change
ARID1A siFOXA1 vs siCont log2 fold change
g
BRG1
MCF7
PCC= - 0.08,
p = 8.23X10-107
BRG1 Tam vs Veh log2 fold change
BRG1 siFOXA1 vs siCont log2 fold change
h
ER binding regions close to ARID1A repressed genes
ARID1A
BRG1
p = 0.0031
p = 2.1X10-12
Normalized tag density
siCont siFOXA1 siCont siFOXA1

Figure 5


a
BRG1 ChIP-seq
b
HDAC1 ChIP-seq
c
ER ChIP-seq
Normalized signal enrichment on HDAC1 lost sites upon ARID1A loss
Distance kBp
WT Tamox
KO-11 Tamox
KO-14 Tamox
d
KO-11 Tamox/WT Tamox
PCC= 0.58,
p = 1X10-16
HDAC1 log2 fold change
BRG1 log2 fold change
e
KO-14 Tamox/WT Tamox
PCC= 0.57,
p = 1X10-16
f
ER qPLEX-RIME
PDX tumors
BAF170
BAF155
HDAC1
WT1 WT2 MT1 MT2
Wild type
ARID1A mutant
Log2 normalised
protein intensities
4 5 6 7

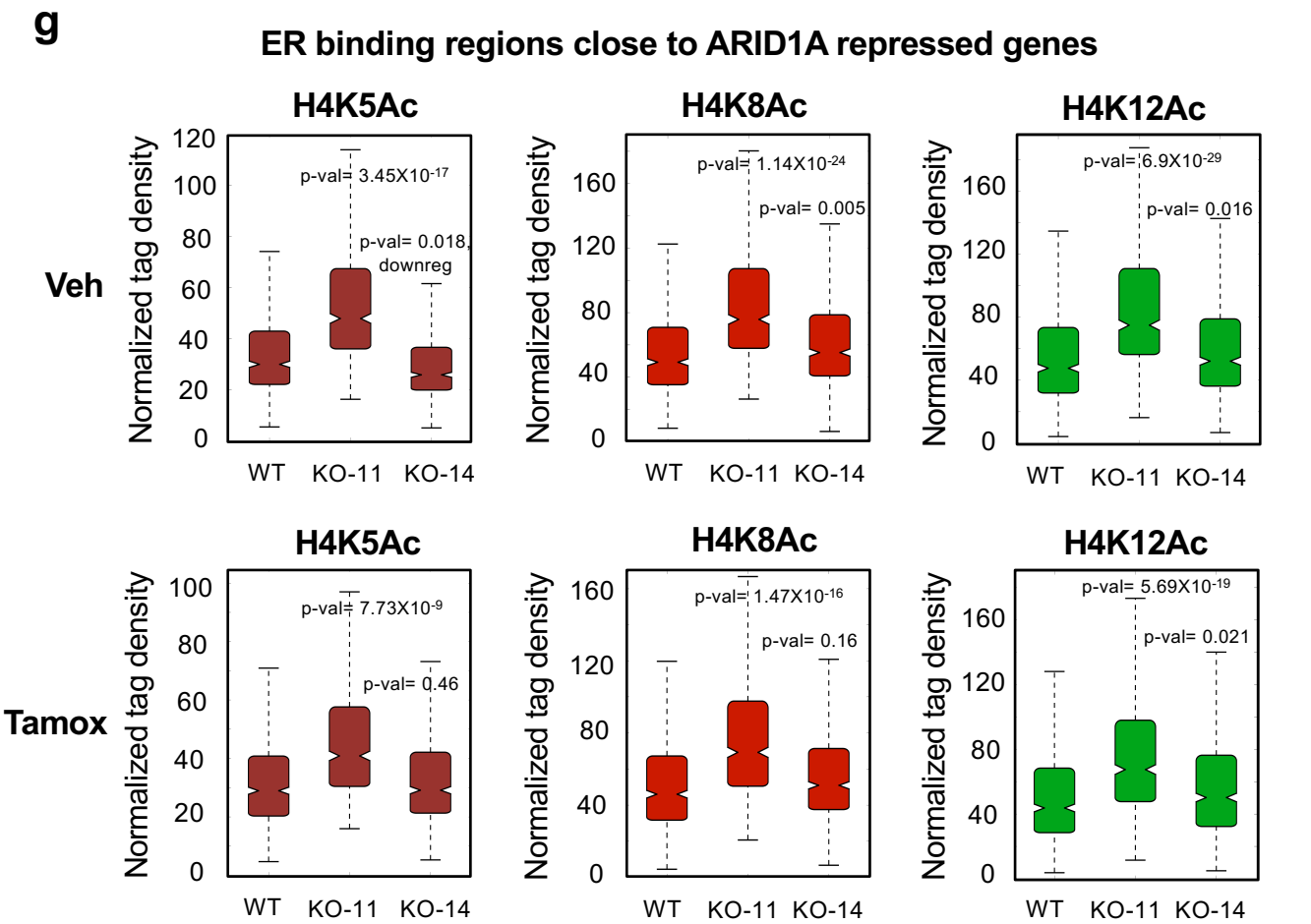
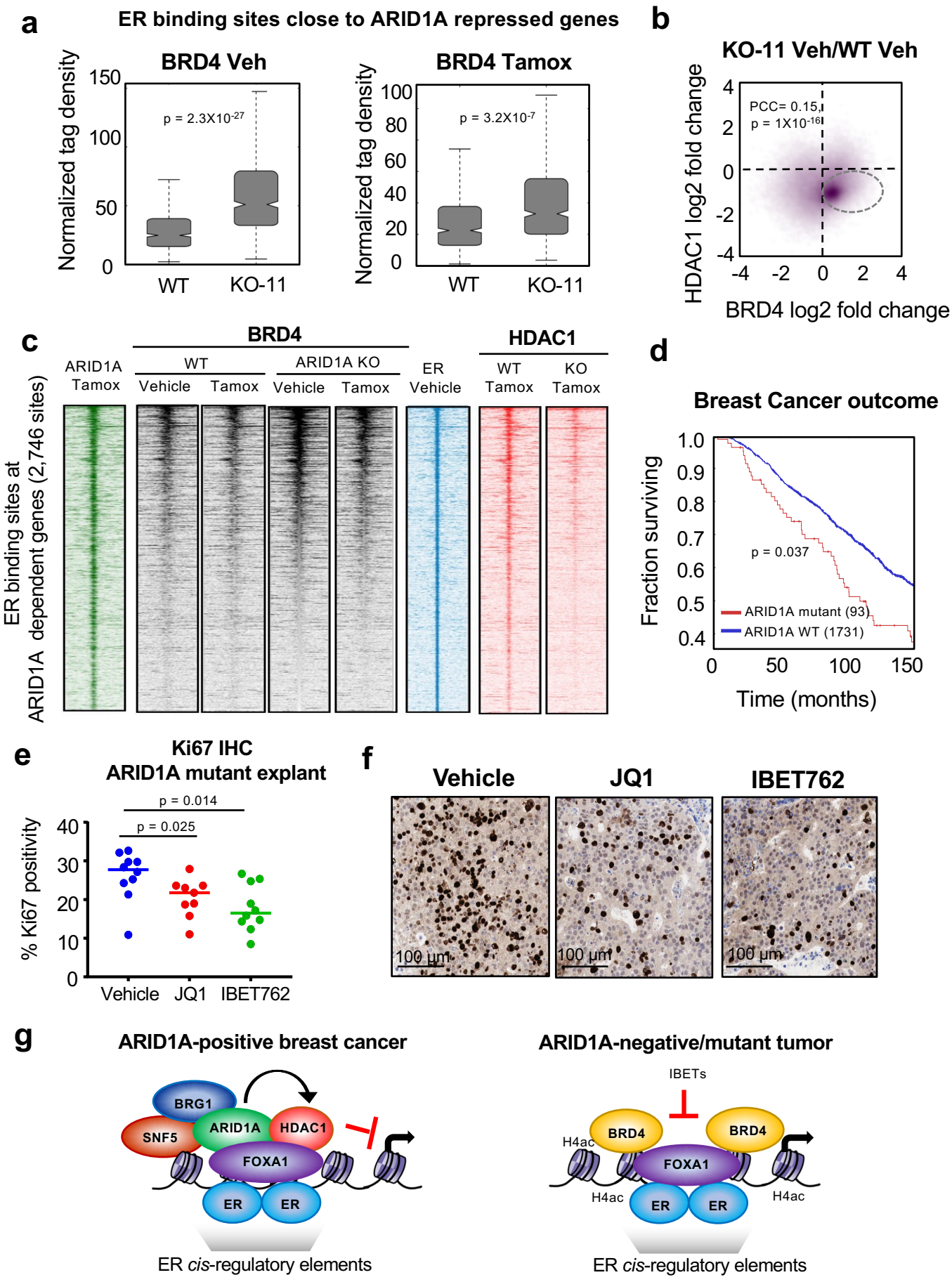
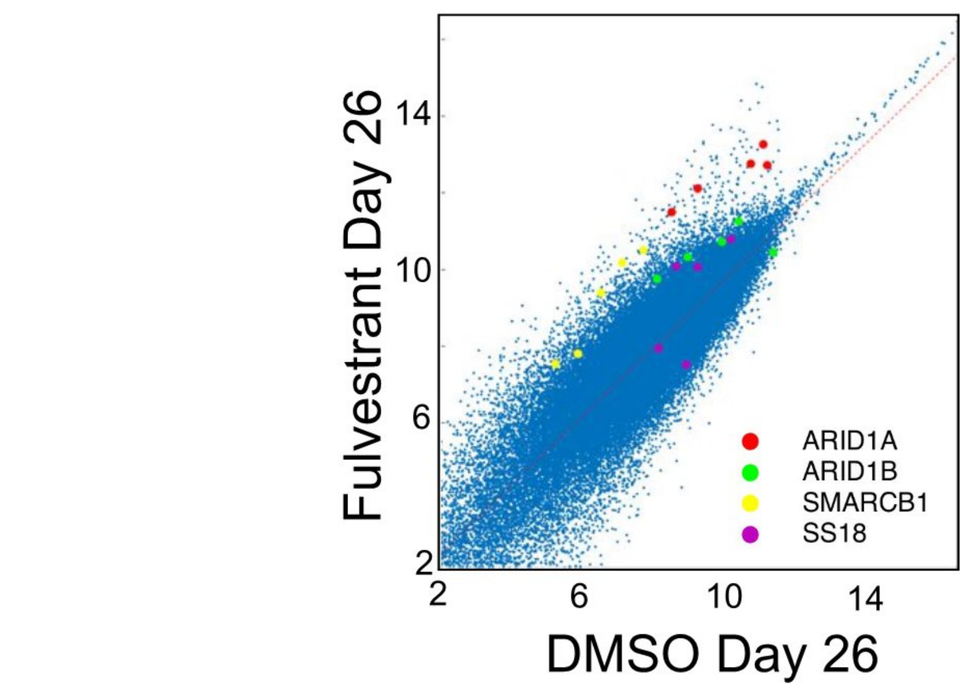
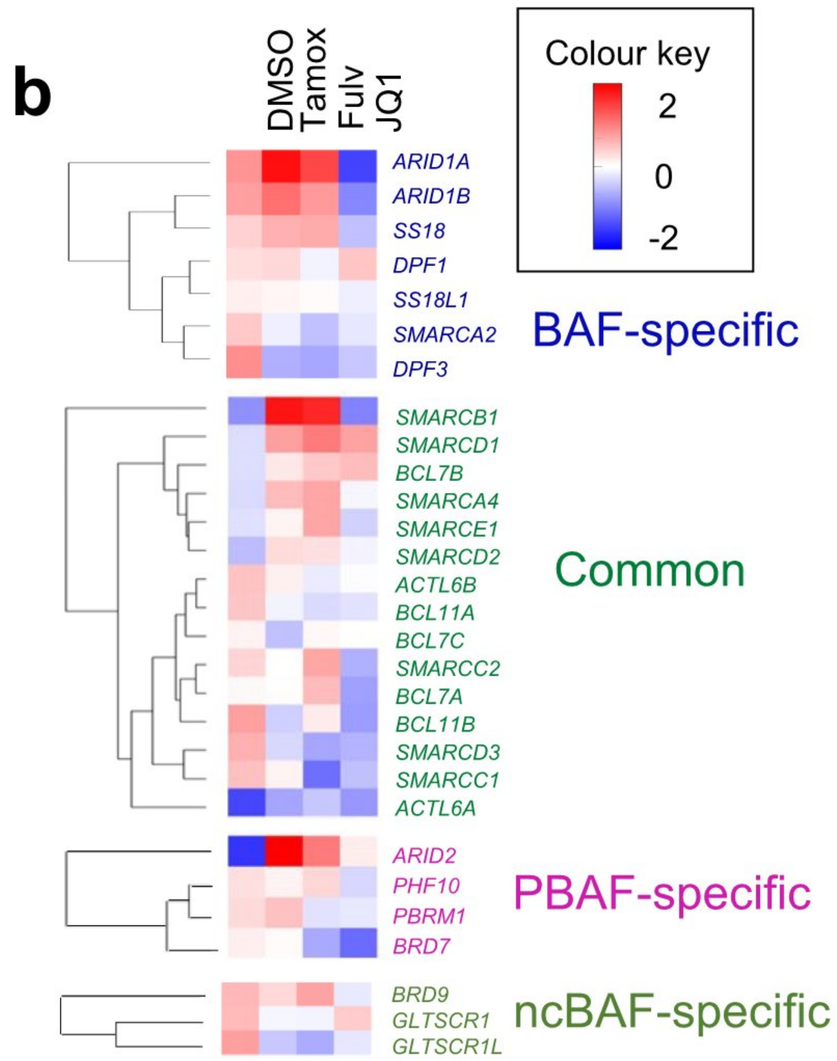

g
ER binding regions close to ARID1A repressed genes
H4K5Ac
H4K8Ac
H4K12Ac
Veh
Tamox
Normalized tag density
p-val= 3.45X10-17
p-val= 0.018,
downreg
p-val= 1.14X10-24
p-val= 0.005
p-val= 6.9X10-29
p-val= 0.016
p-val= 7.73X10-9
p-val= 0.46
p-val= 1.47X10-16
p-val= 0.16
p-val= 5.69X10-19
p-val= 0.021
WT KO-11 KO-14

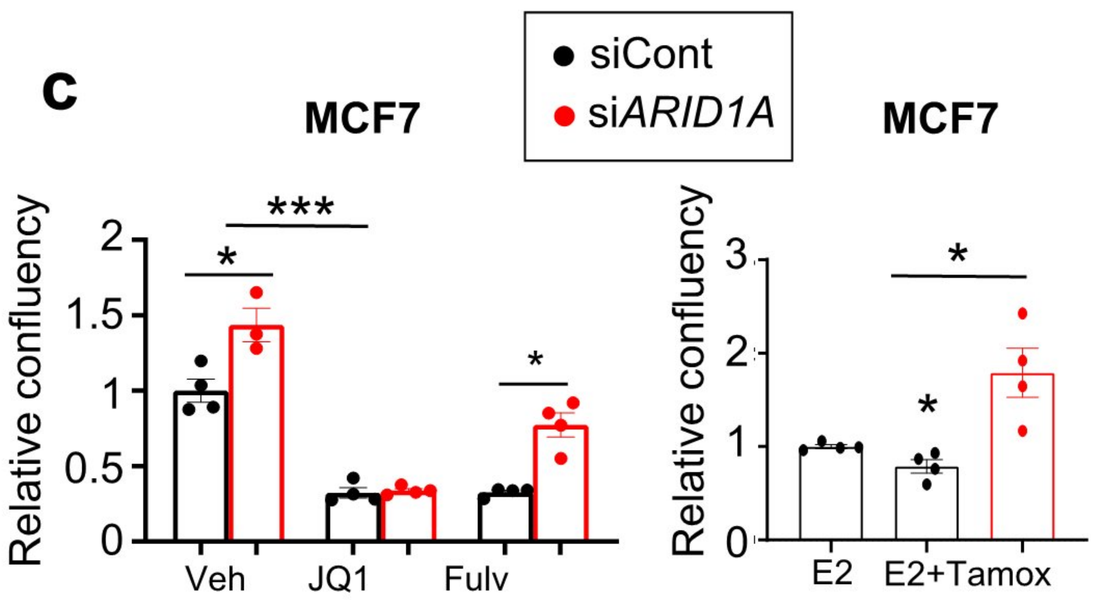
Figure 6

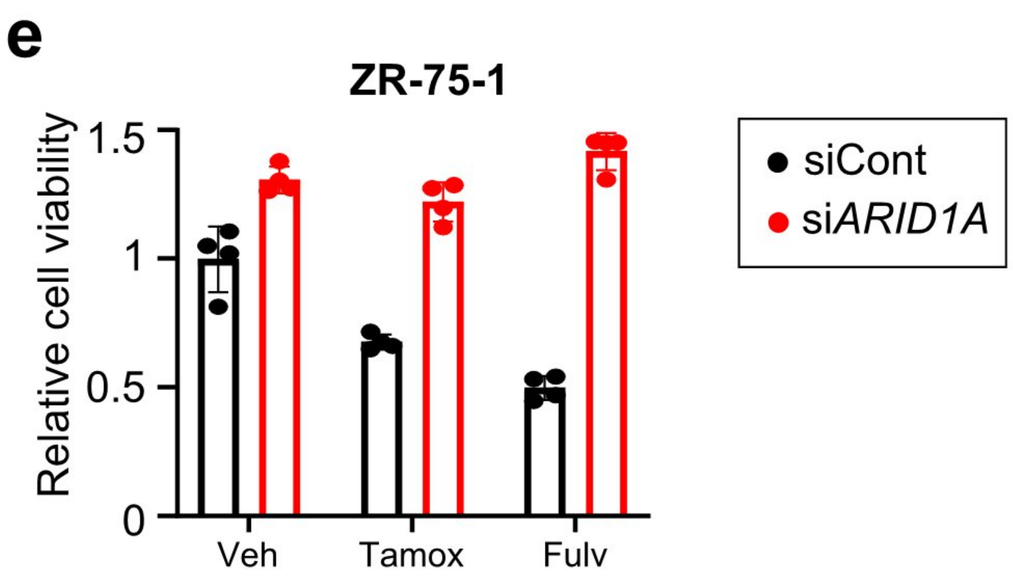

a
ER binding sites close to ARID1A repressed genes
BRD4 Veh
Normalized tag density
p = 2.3X10-27
WT
KO-11
BRD4 Tamox
p = 3.2X10-7
b
KO-11 Veh/WT Veh
HDAC1 log2 fold change
PCC= 0.15
p = 1X10-16
BRD4 log2 fold change
c
BRD4
HDAC1
ARID1A Tamox
WT
ARID1A KO
Vehicle
Tamox
ER Vehicle
KO Tamox
ER binding sites at ARID1A dependent genes (2,746 sites)
d
Breast Cancer outcome
Fraction surviving
p = 0.037
ARID1A mutant (93)
ARID1A WT (1731)
Time (months)
e
Ki67 IHC
ARID1A mutant explant
% Ki67 positivity
p = 0.014
p = 0.025
Vehicle
JQ1
IBET762
f
Vehicle
JQ1
IBET762
100 µm
g
ARID1A-positive breast cancer
BRG1
SNF5
ARID1A
HDAC1
FOXA1
ER
ER cis-regulatory elements
ARID1A-negative/mutant tumor
IBETs
BRD4
H4ac
FOXA1
ER cis-regulatory elements

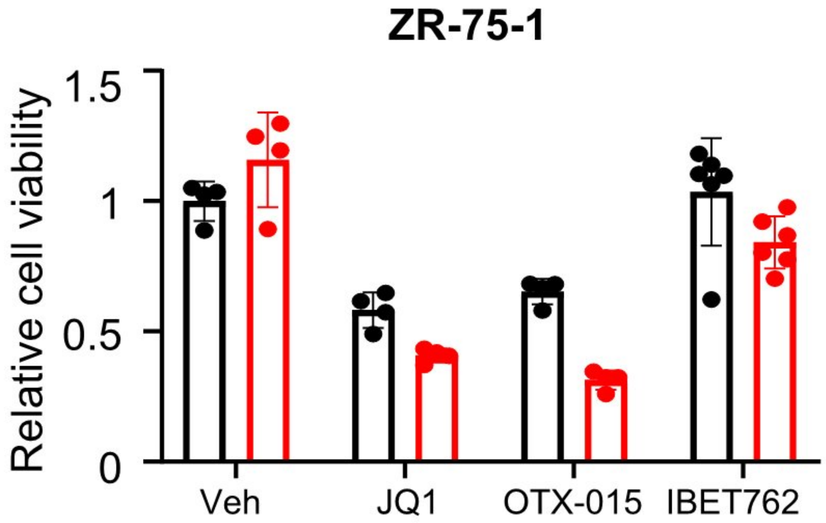
Extended Data Fig. 1



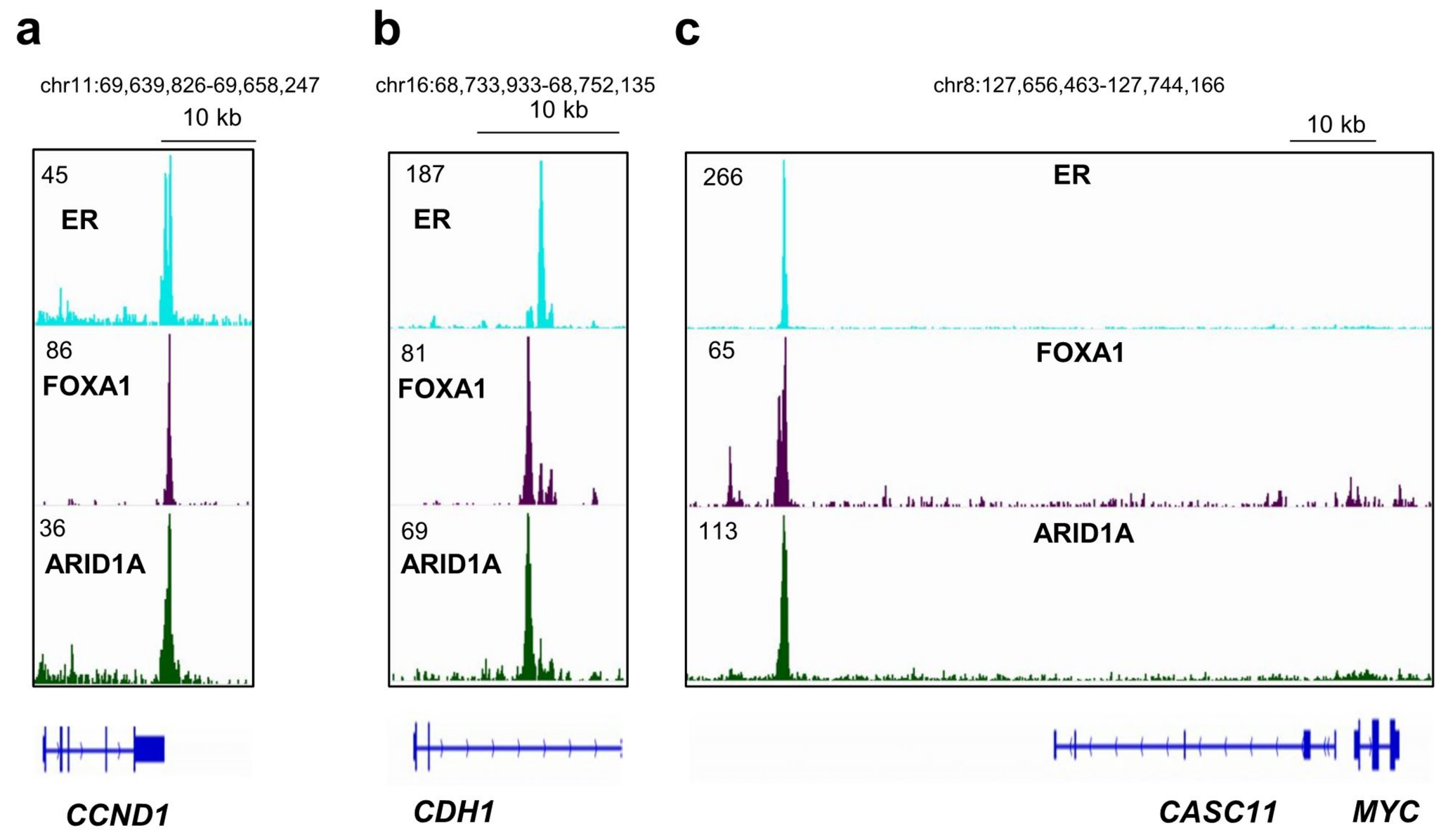


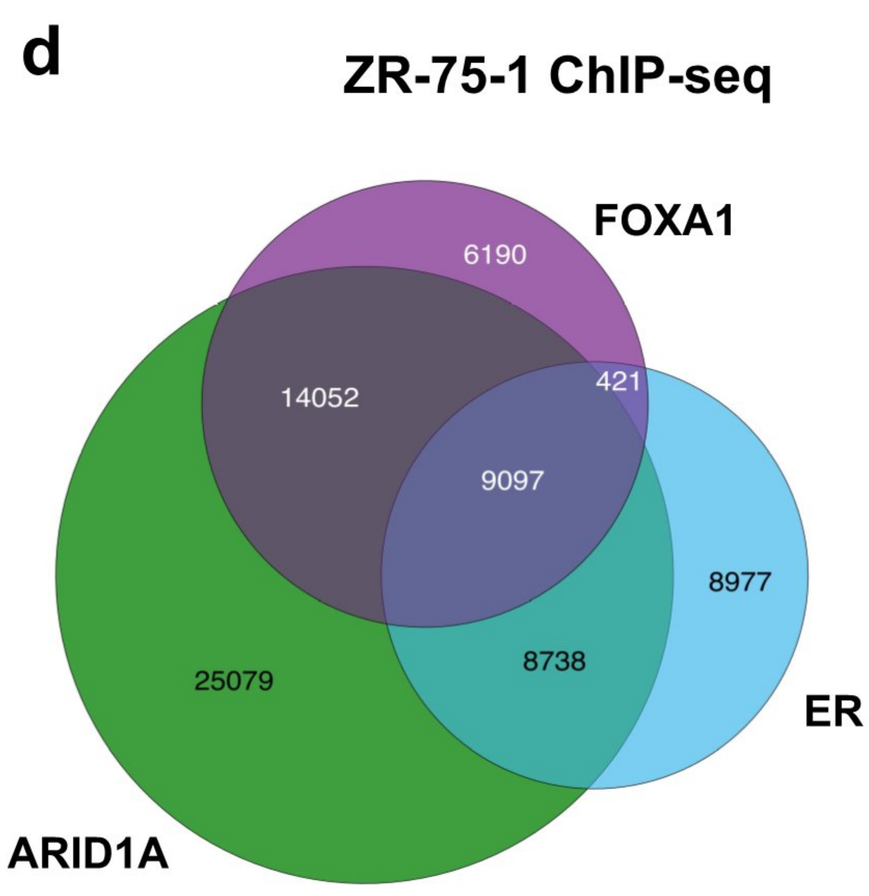


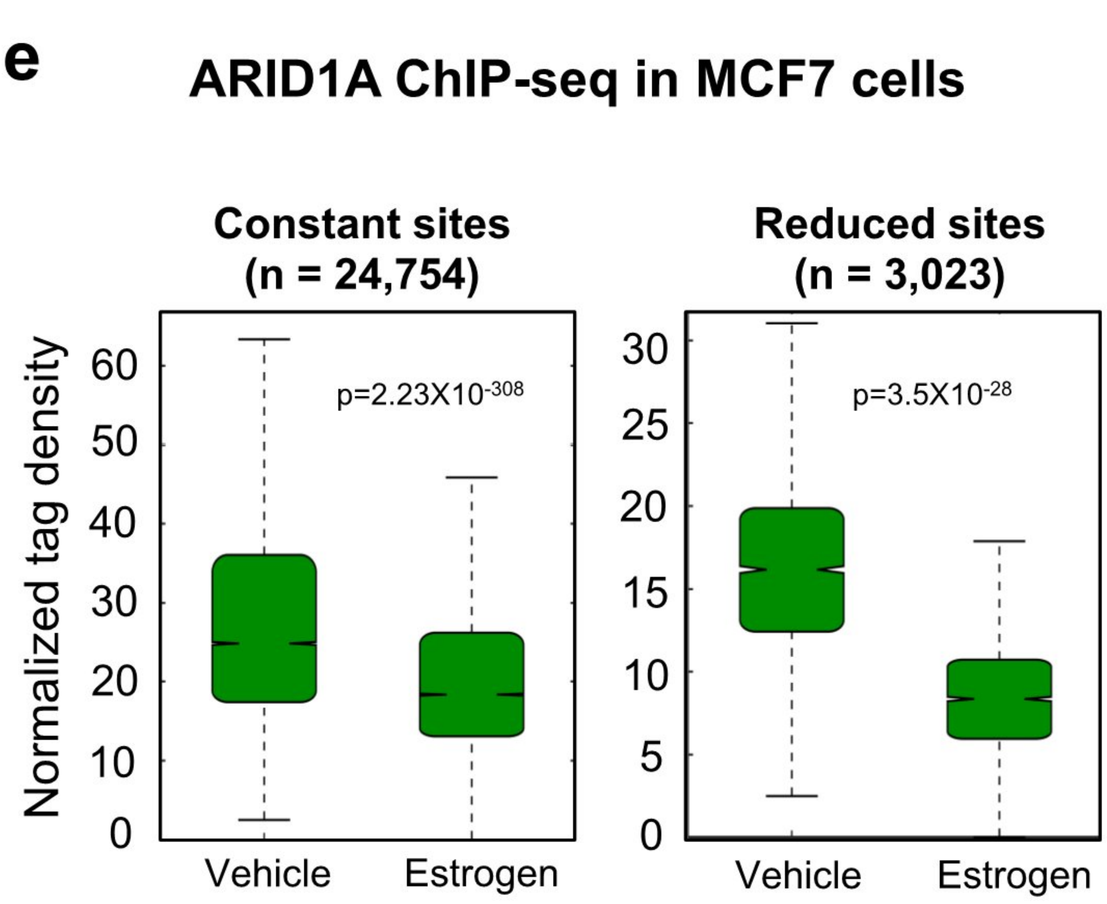




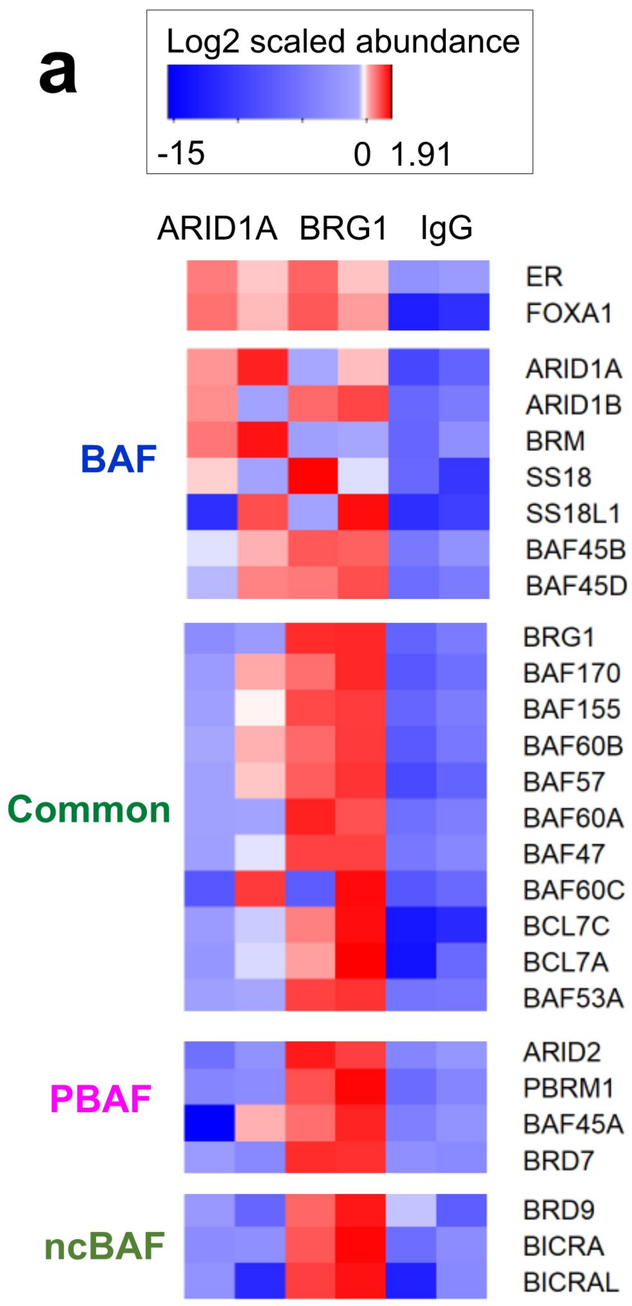
Extended Data Fig. 2

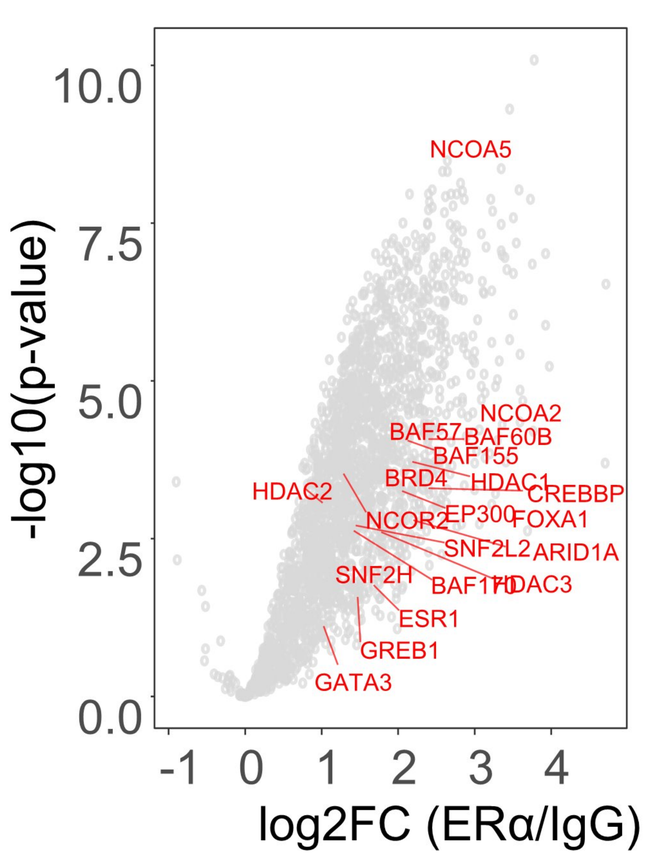


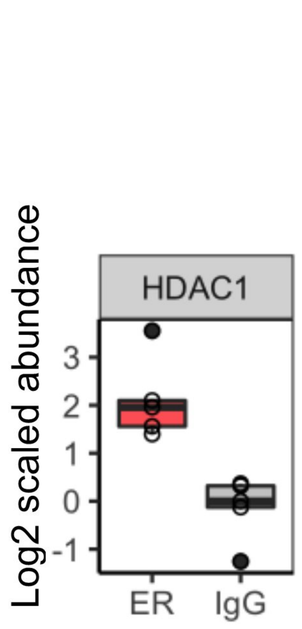


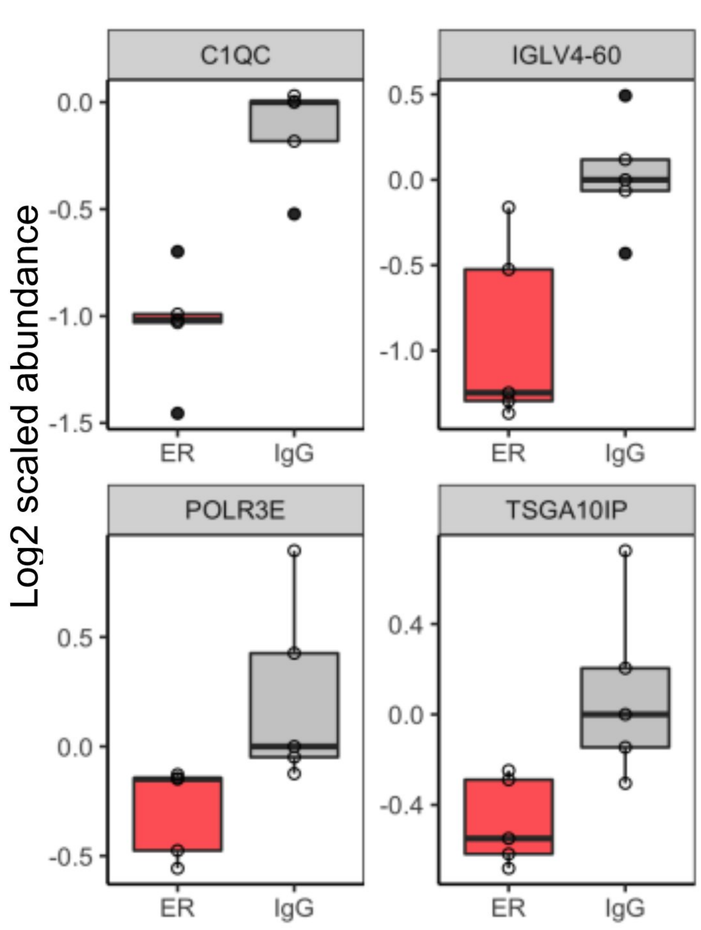


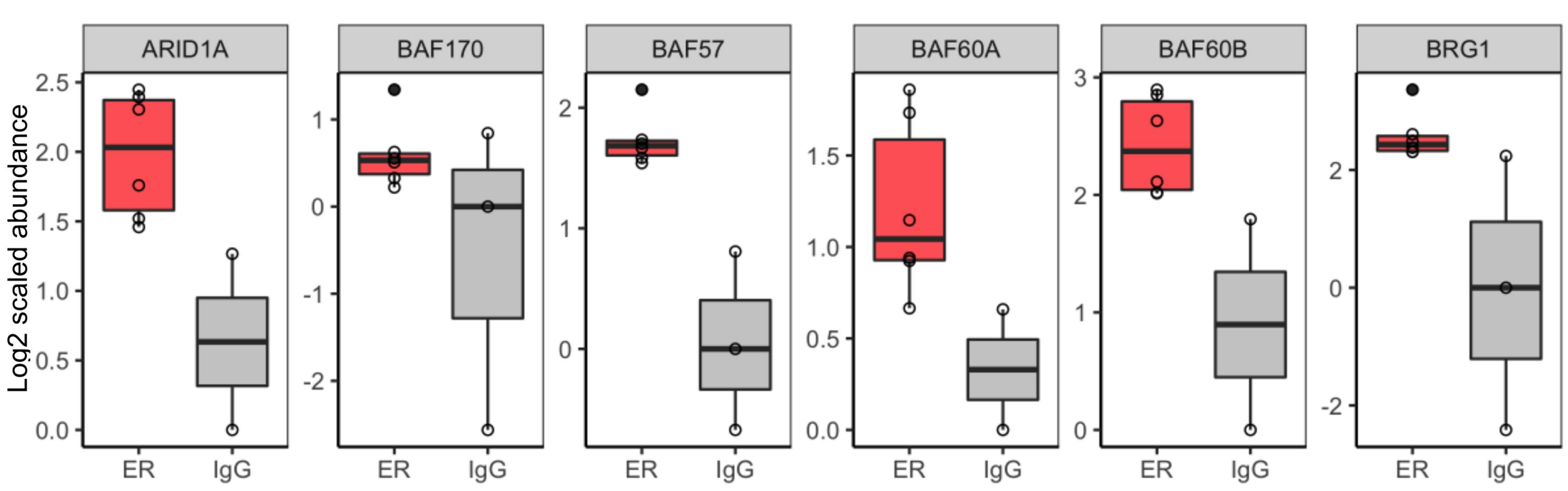
Extended Data Fig. 3



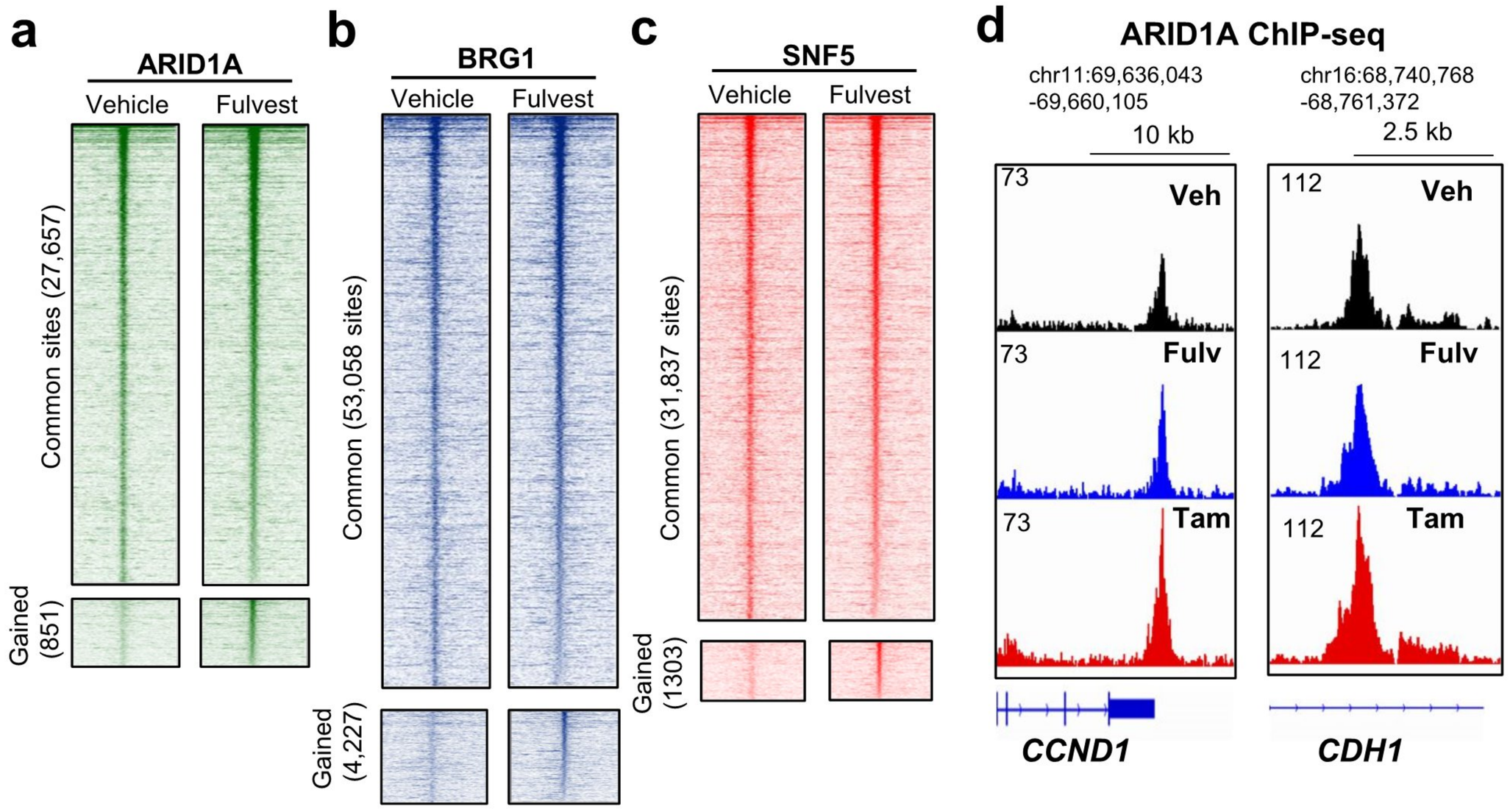


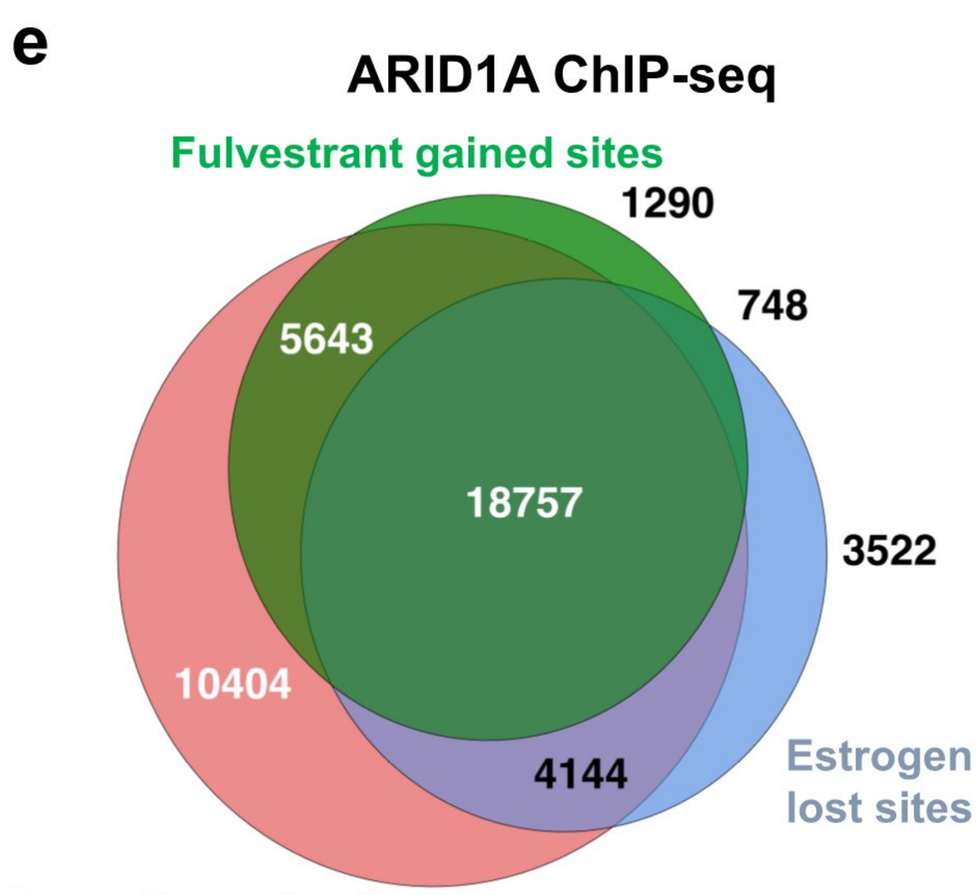


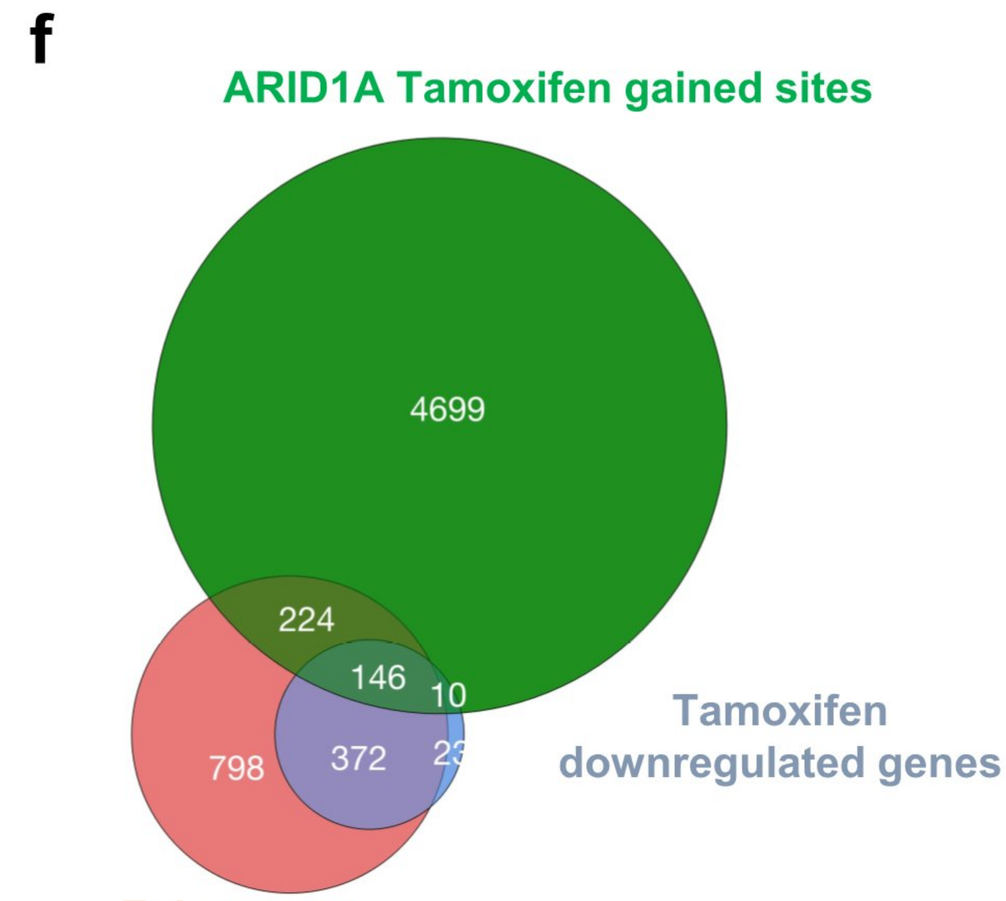




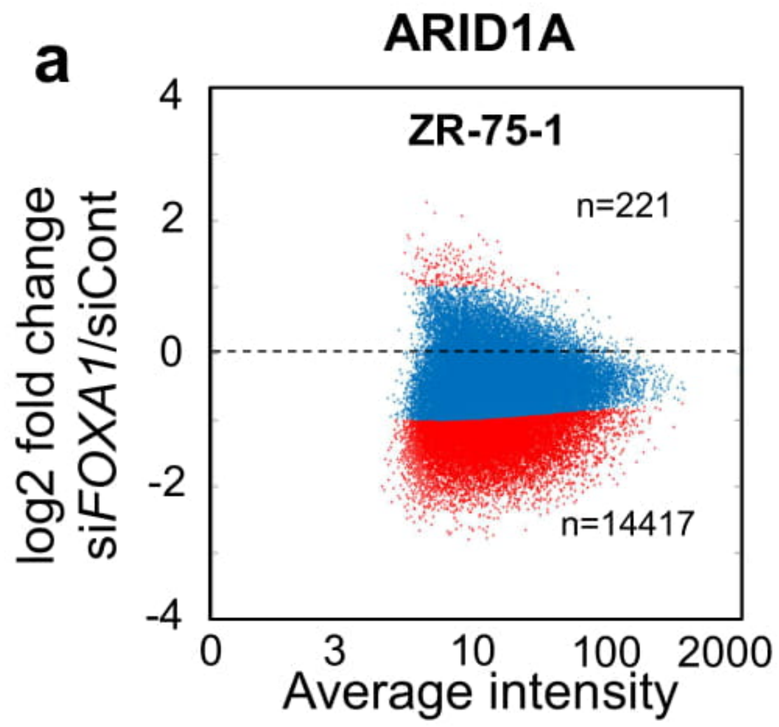
Extended Data Fig. 4

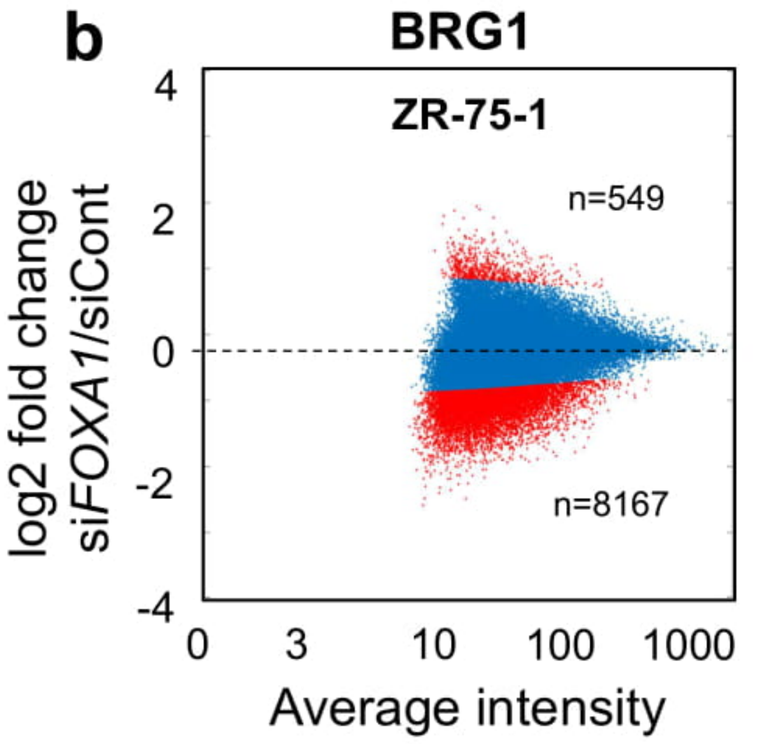


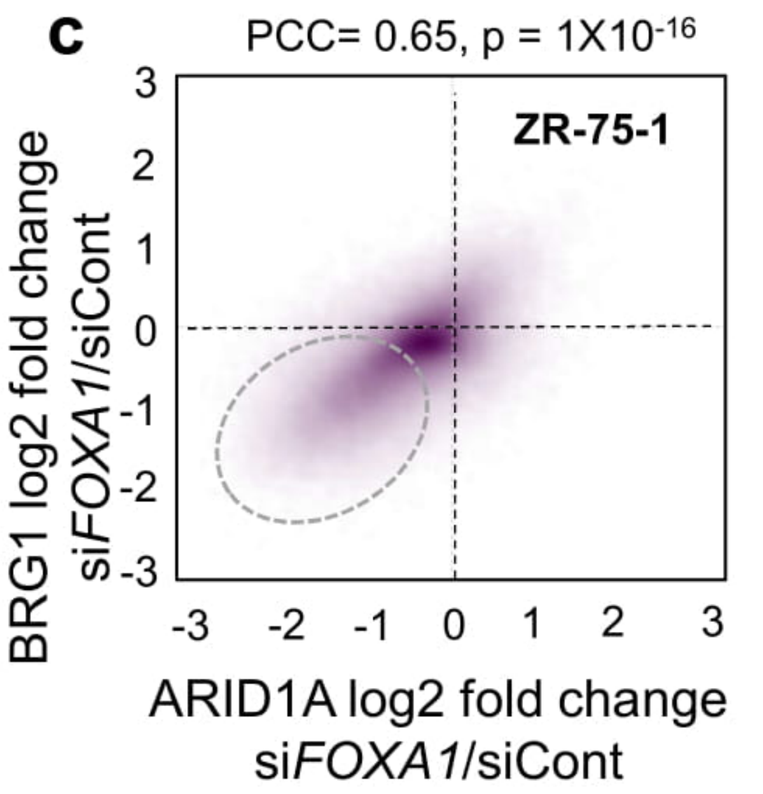


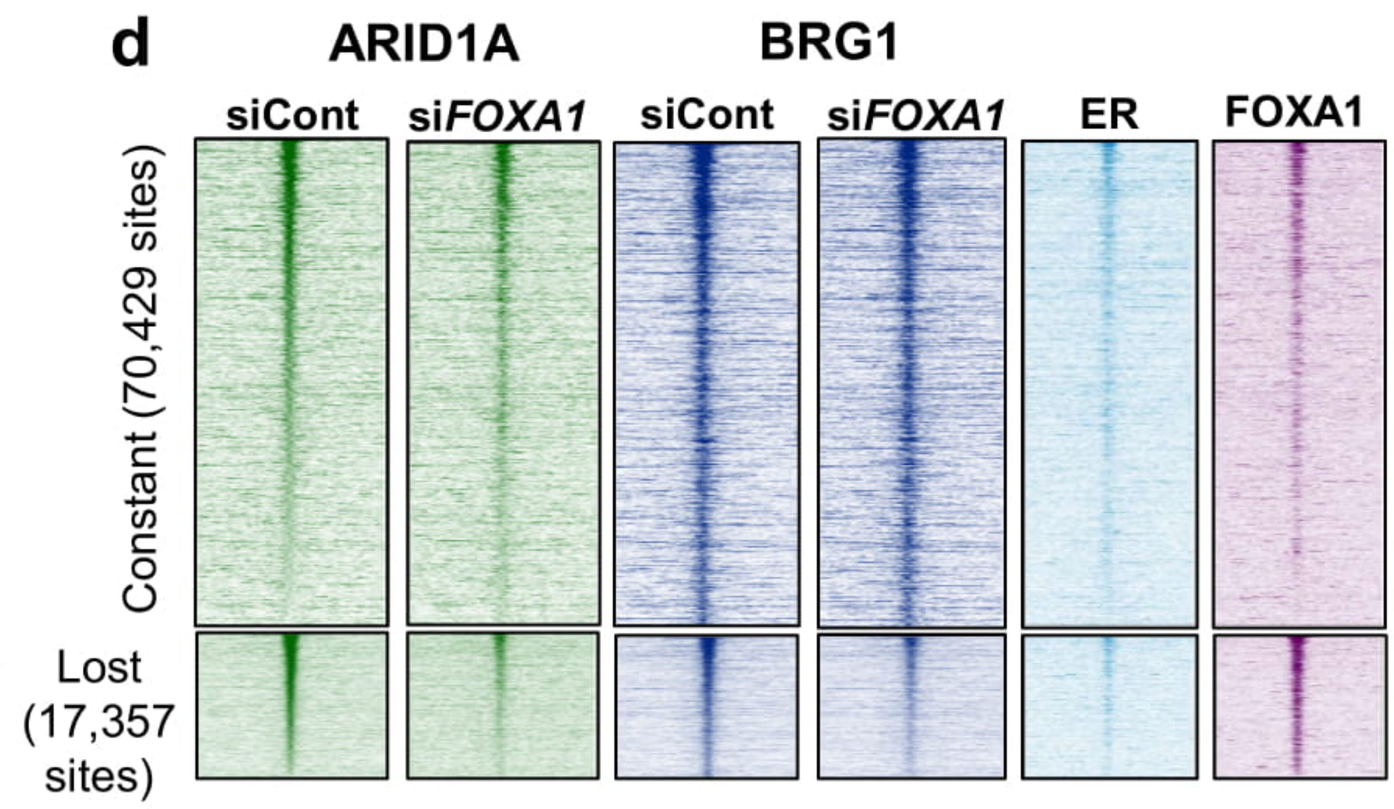


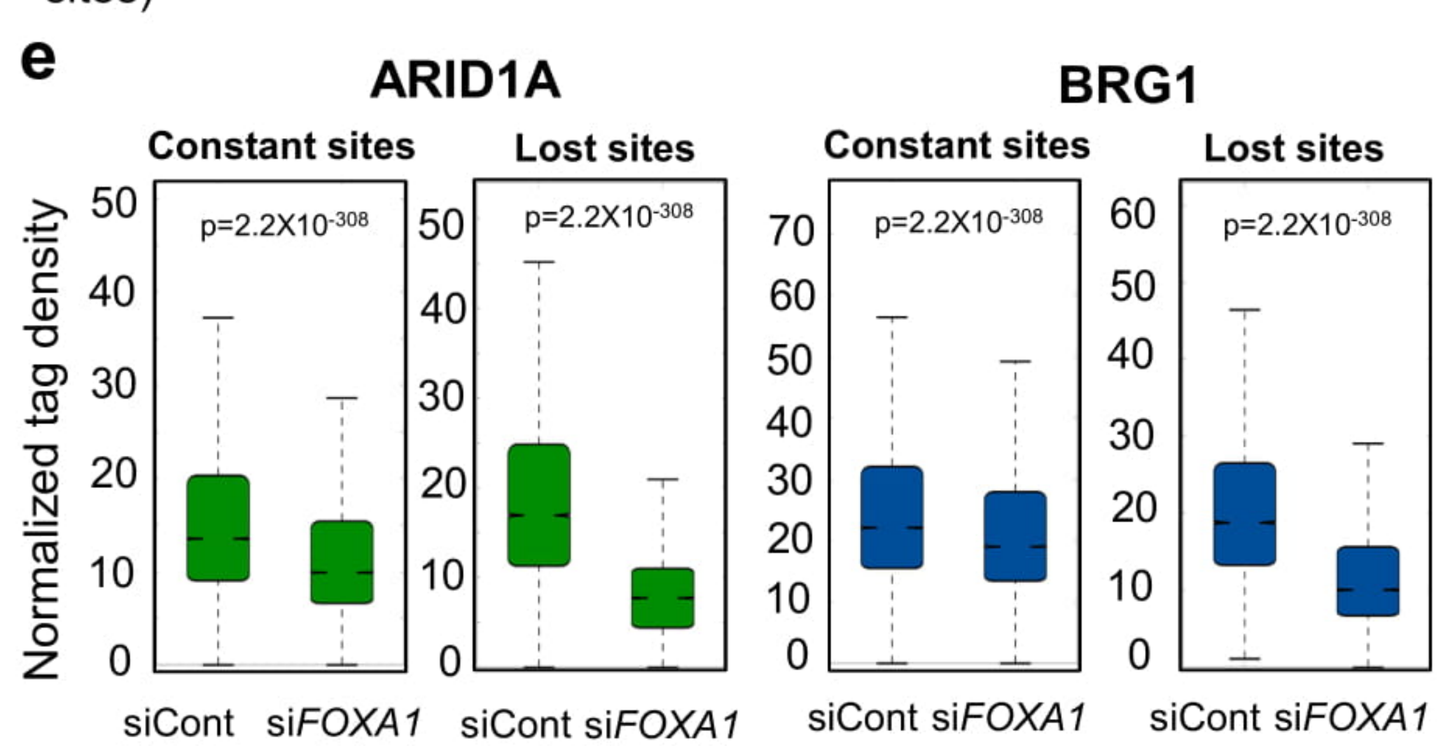
Extended Data Fig. 5



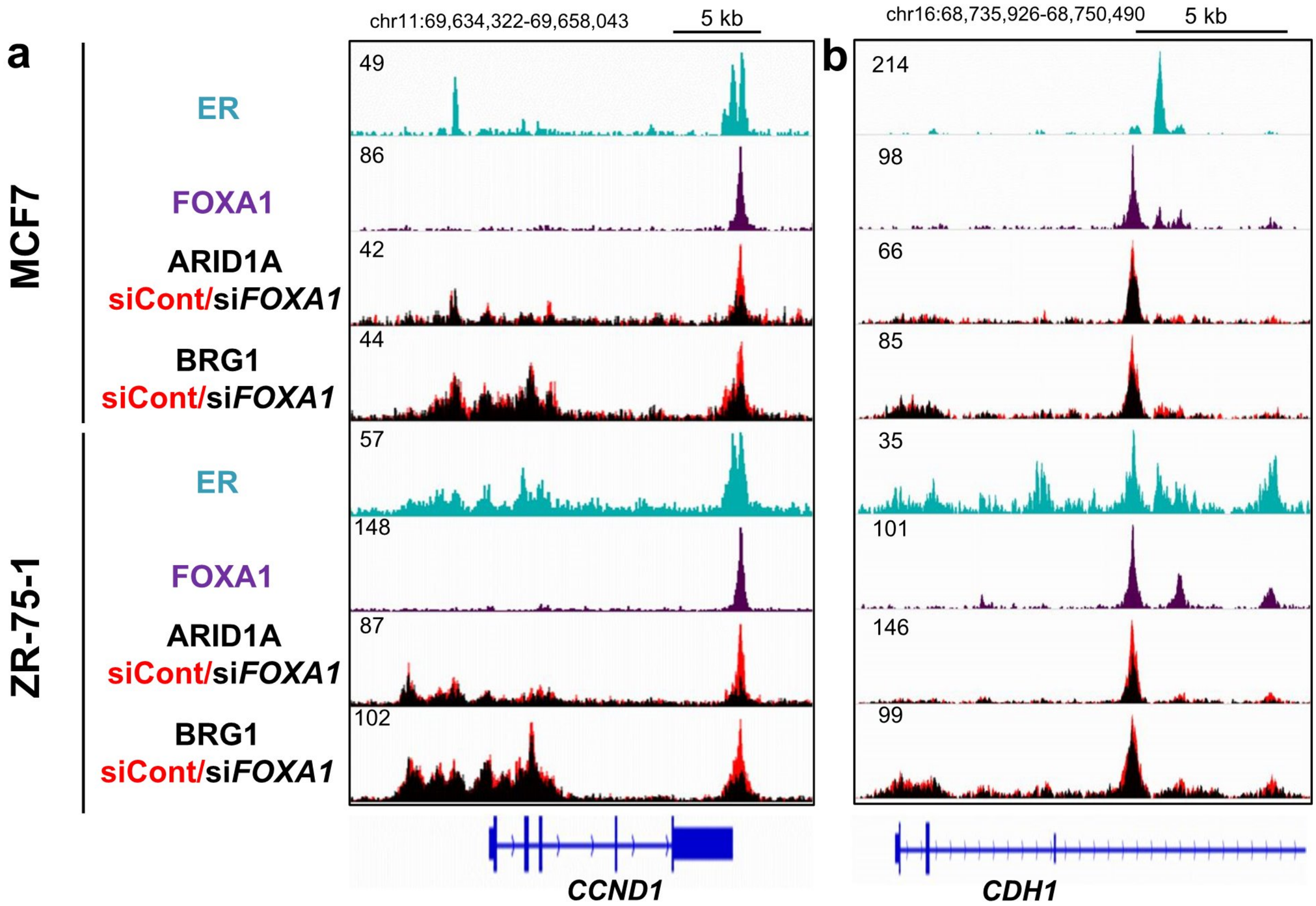


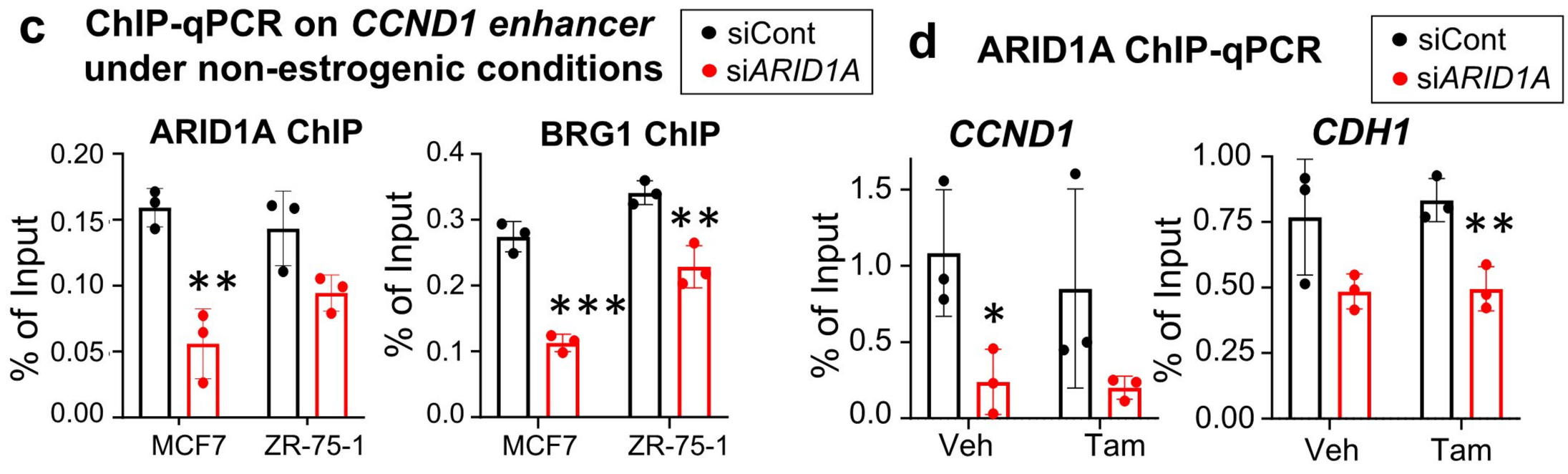






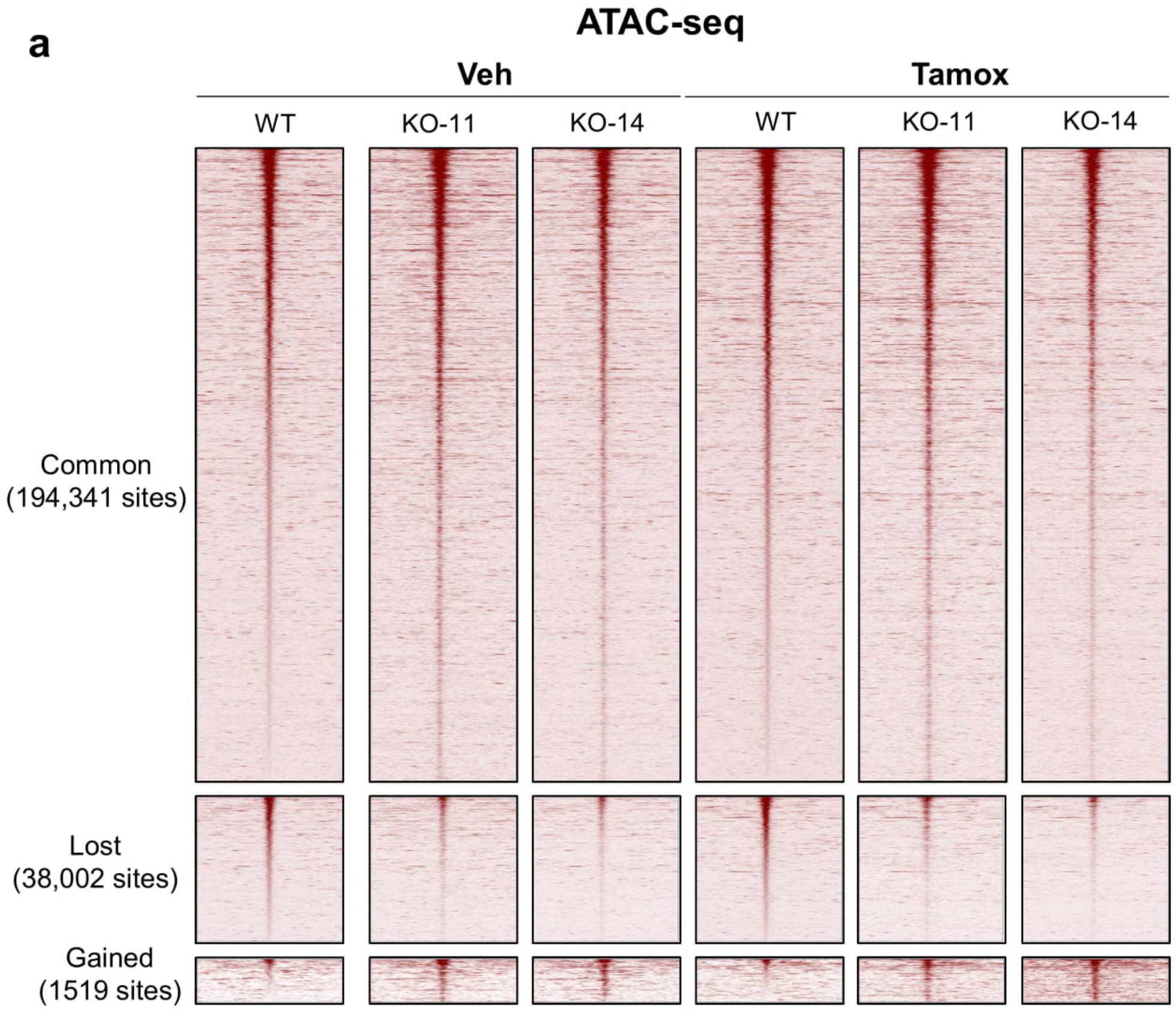
Extended Data Fig. 6



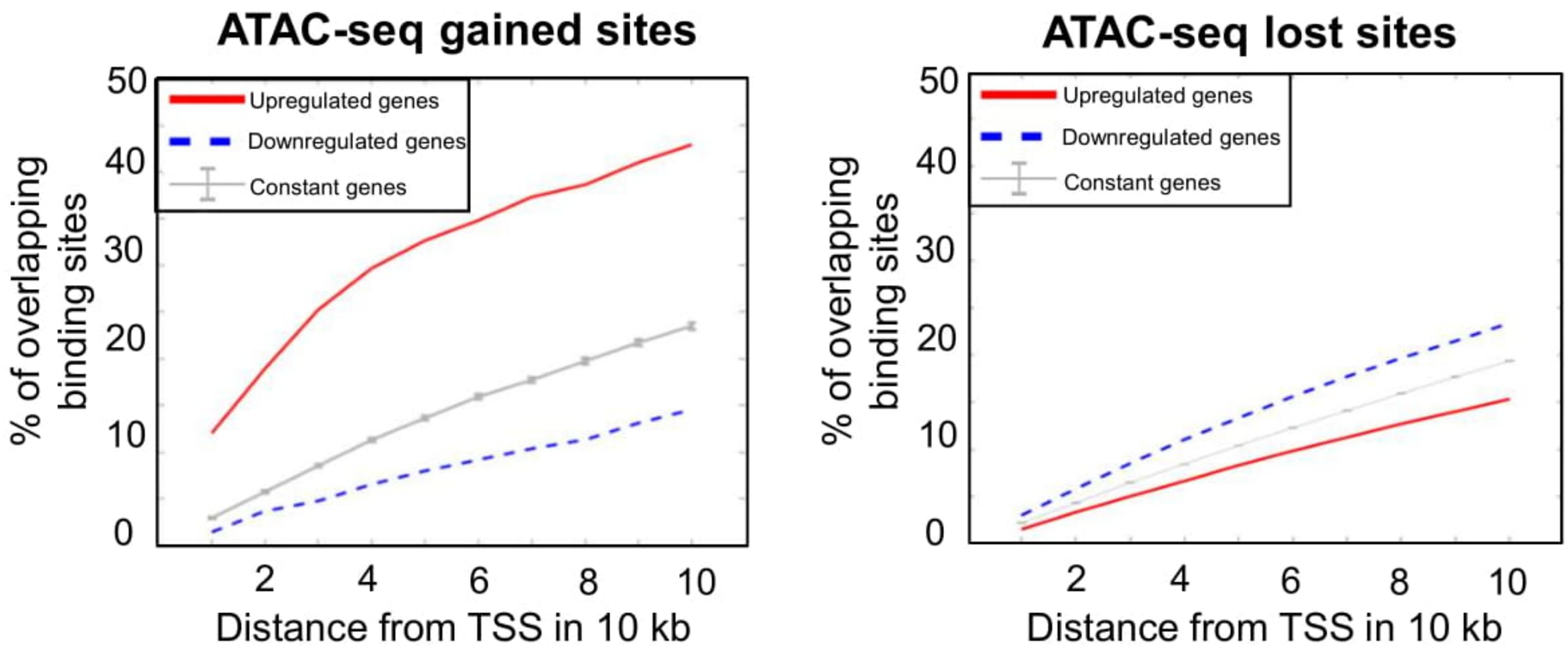


Extended Data Fig. 7

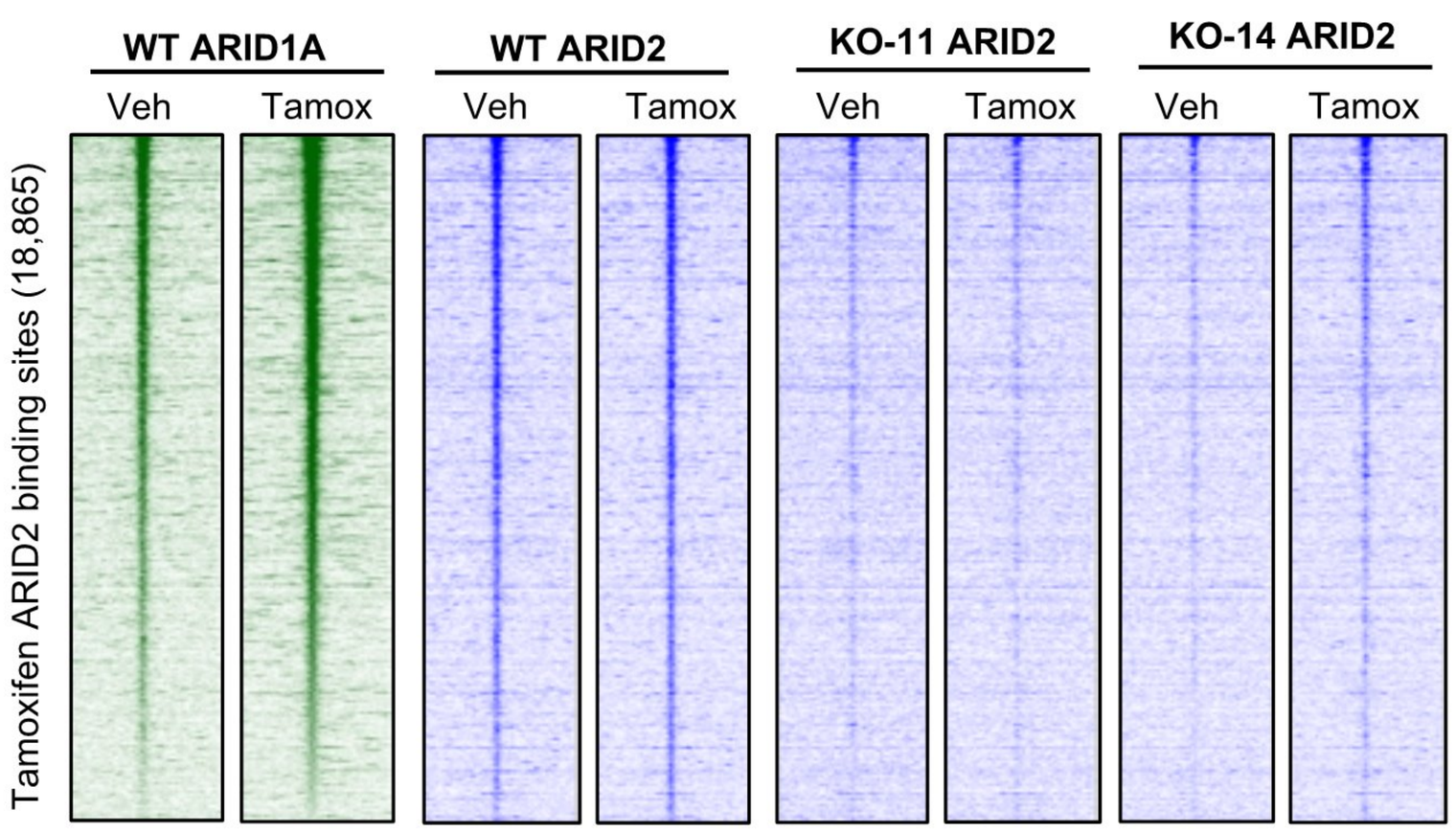
a

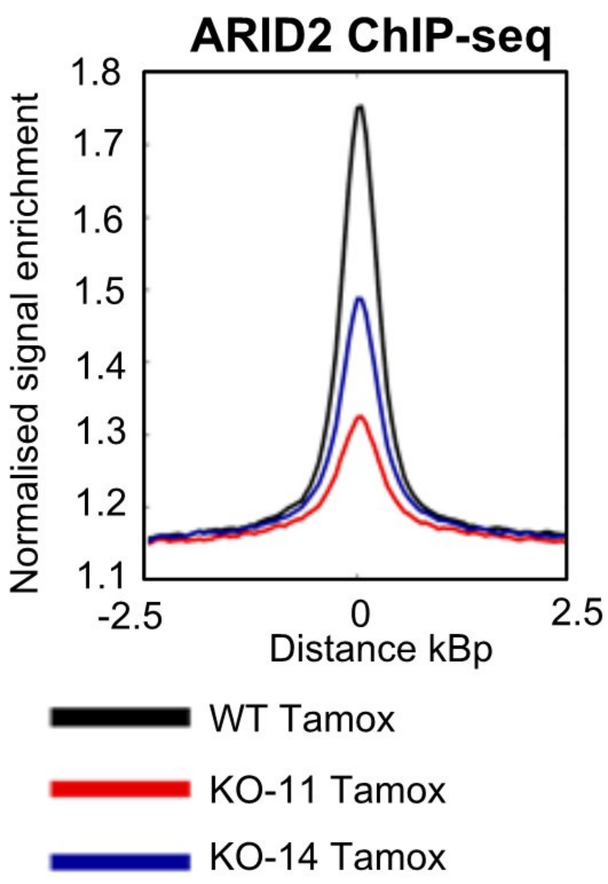


b

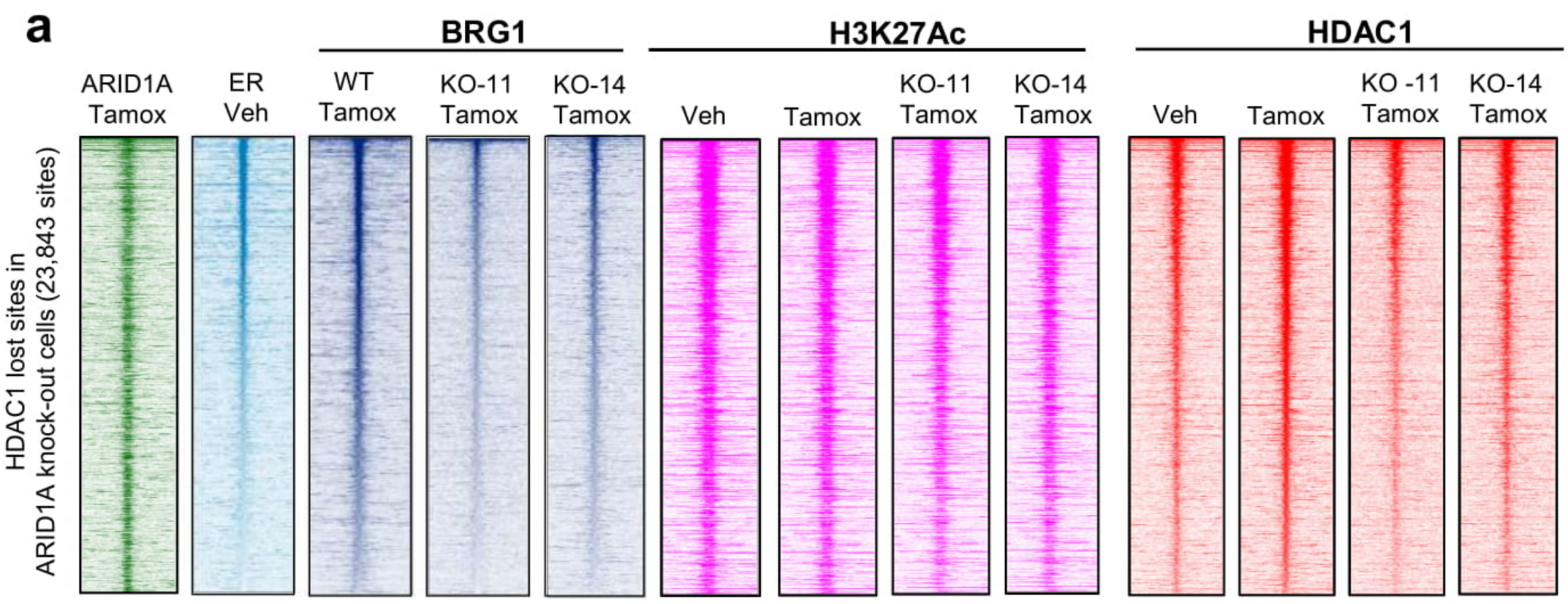


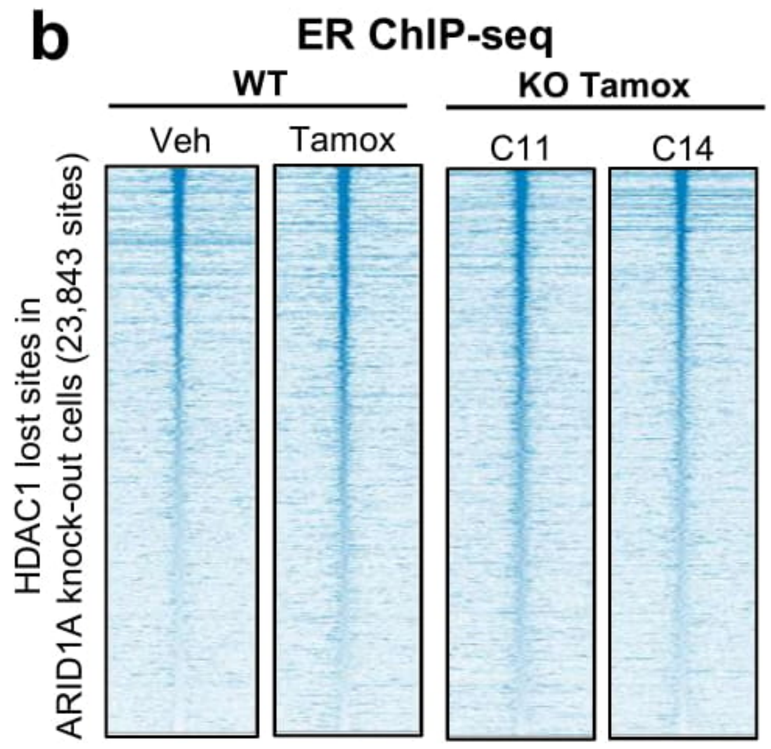
Extended Data Fig. 8

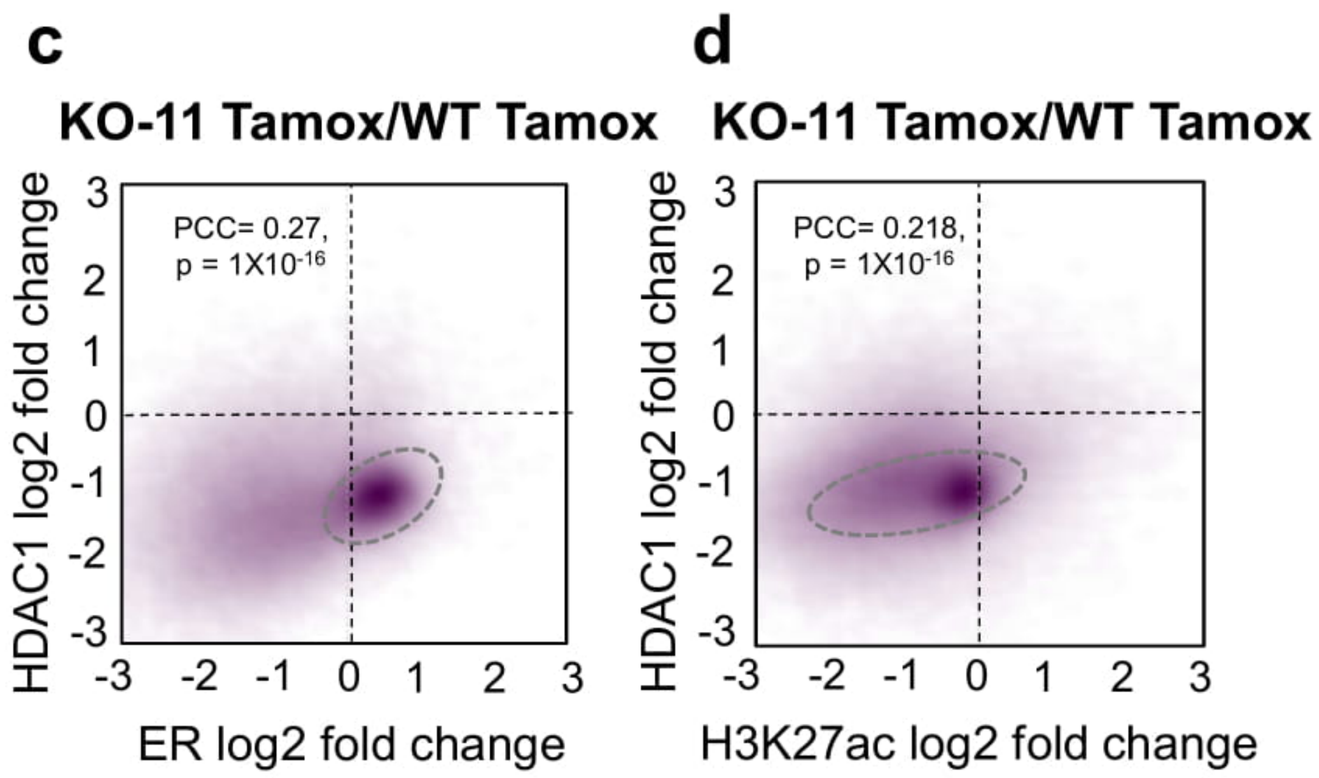


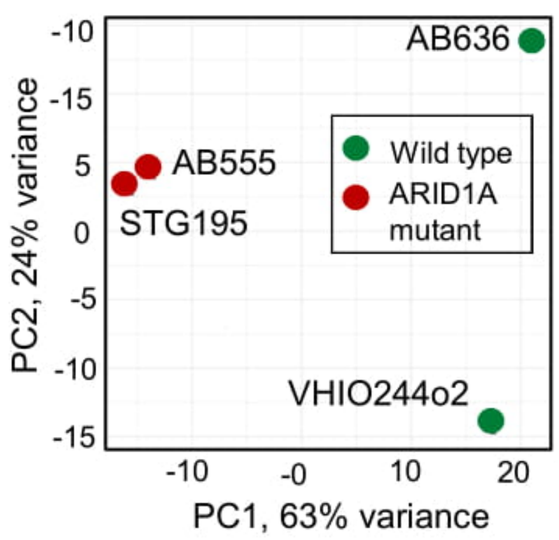


Extended Data Fig. 9





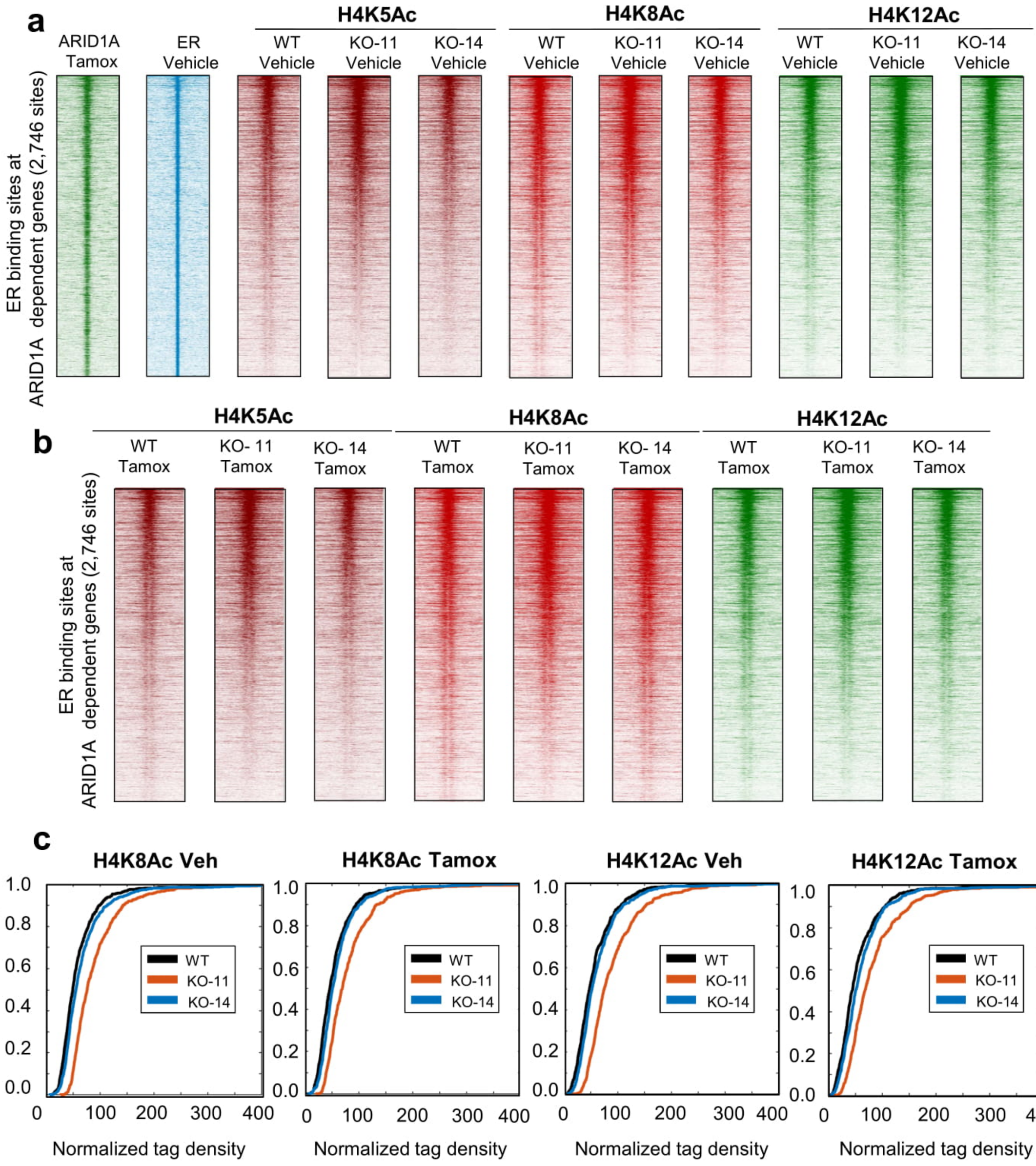




f

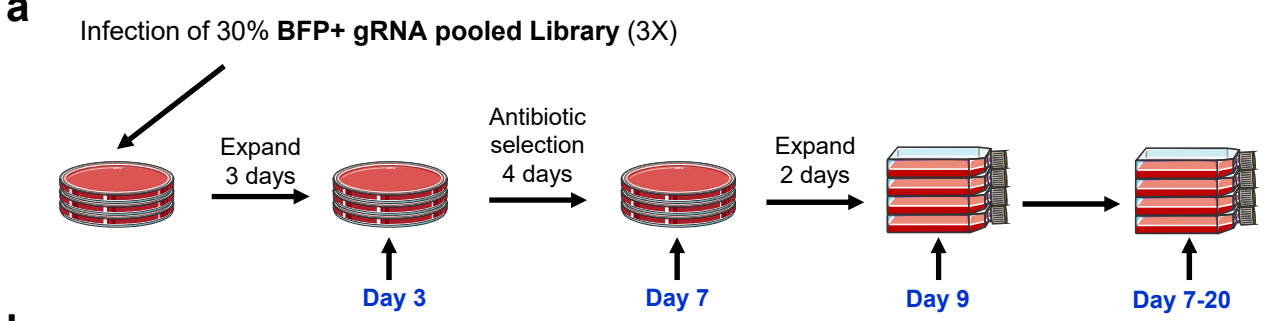
Gene	Model	Mutation	Chromosome	Position	Reference allele	Alternate allele	Type
ARID1A	AB555	Yes	Chr 1	27107135	C	CA	Frameshift
ARID1A	STG195	Yes	Chr 1 Chr 1	27099098 27105761	C C	T T	Stop gained Non-synonymous

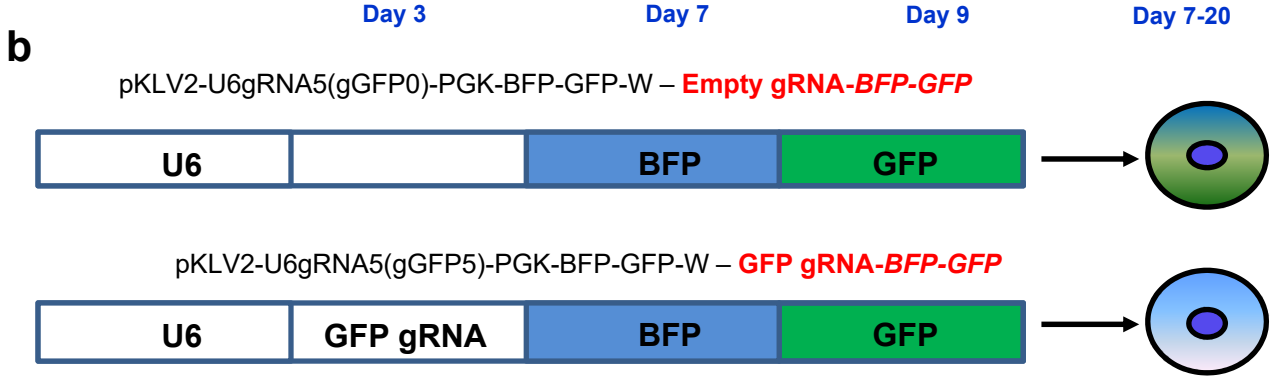
Extended Data Fig. 10

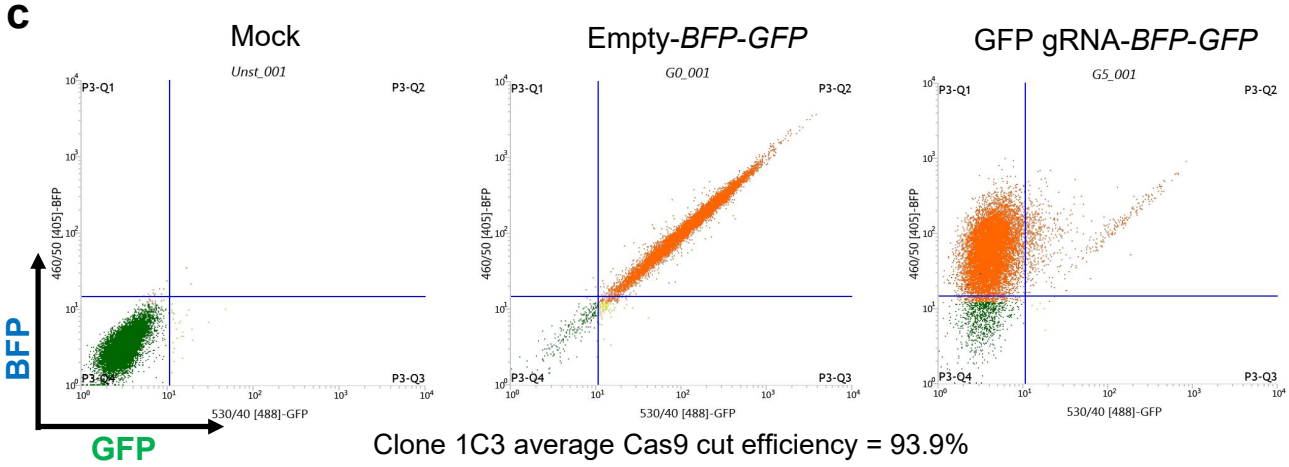


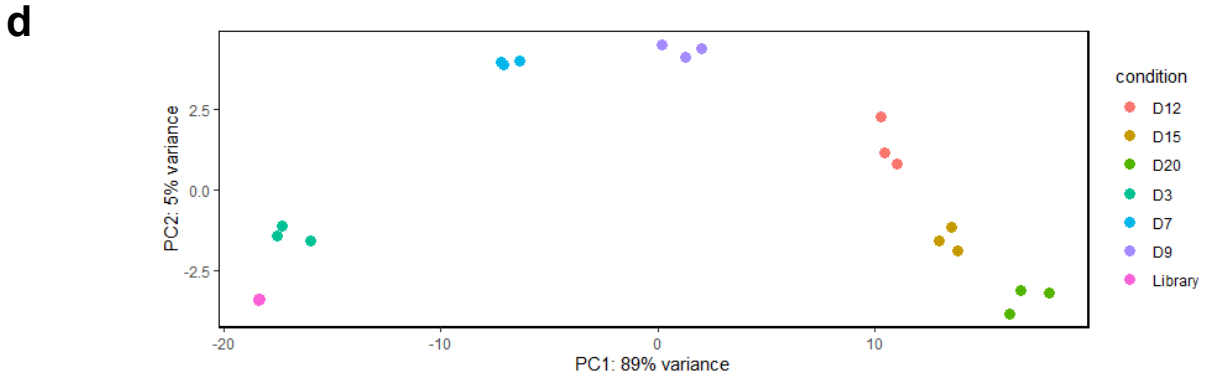
Supplementary Figure 1

Genome-wide CRISPR screening workflow and clone validation of Cas9 expressing cells **a**. CRISPR screening experimental plan to find essential genes from different number of days (in blue) post-infection. **b**. Cas9 efficiency test reporter plasmids used for validating efficiency of the system. **c.** Efficient Cas9-expressing MCF7 clone 1C3 expresses only BFP intensity (y-axis) as the gRNA for GFP has deleted GFP expression (x-axis). n= 4 independent experiments and a representation is shown. **d.** Principal Component analysis (PCA) of the CRISPR screening with different days post-infection with three independent biological cell cultures per time point. n = 3 independent biological cell culture samples.



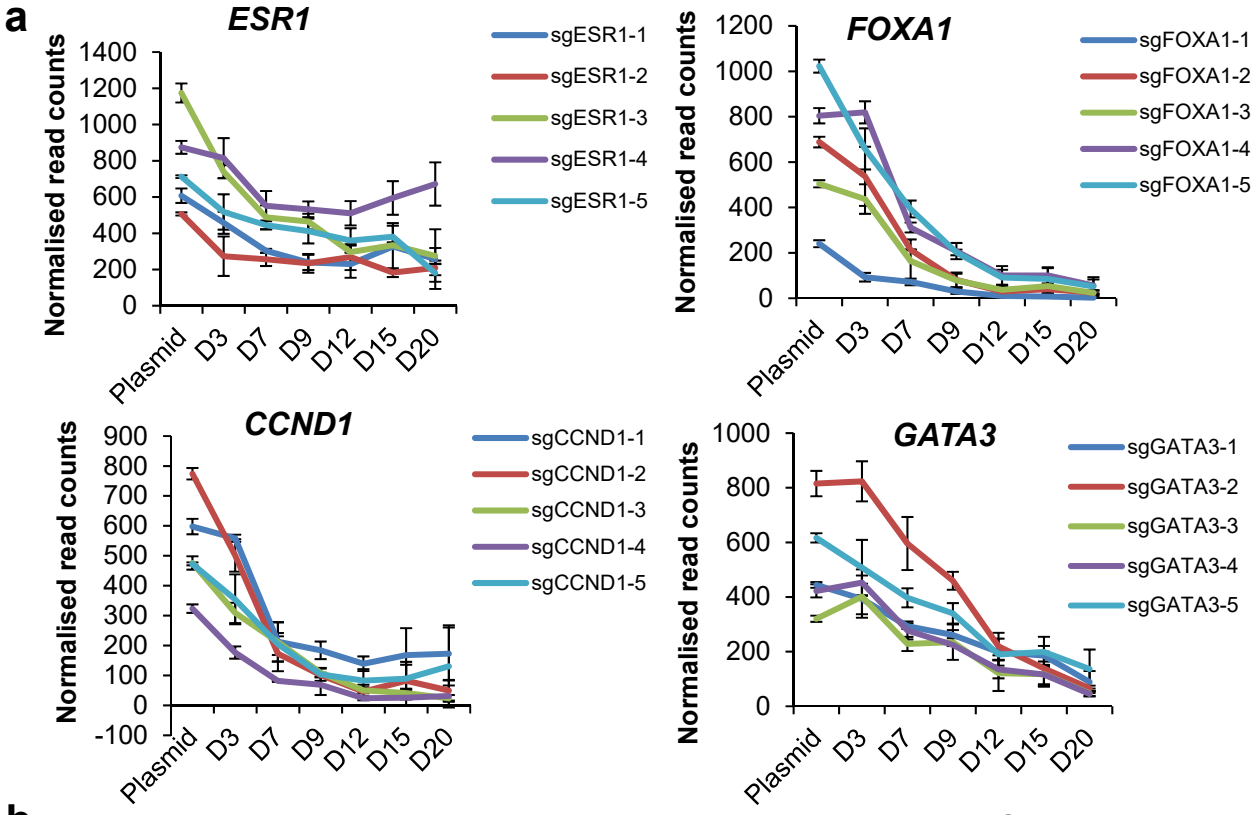


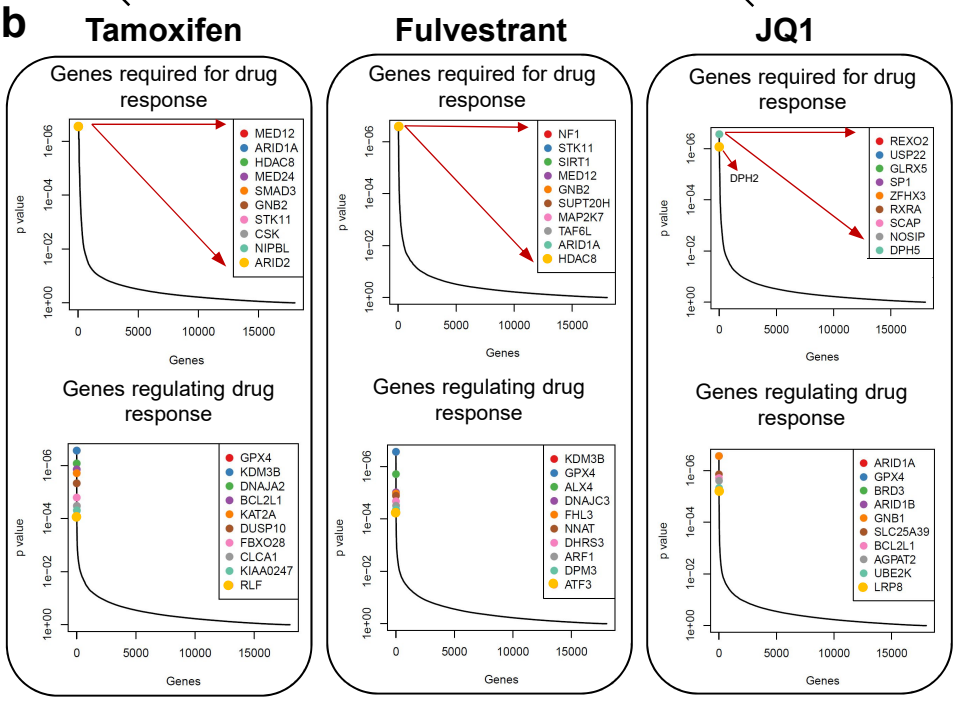


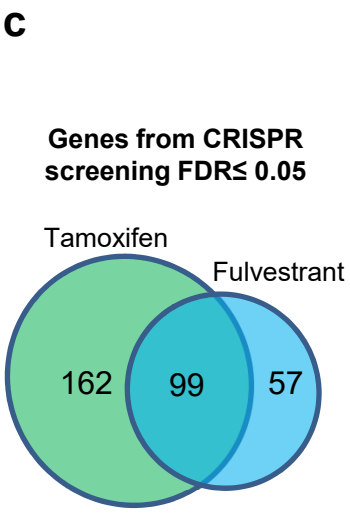


Supplementary Figure 2

Examples of top gene lists from CRISPR screen. **a**. Plots showing changes in individual gRNAs for known ER associated genes or target genes in untreated conditions comparing to uninfected gRNA pool (Plasmid). n=3 independent biological cell cultures, mean ±SD. **b.** Top 10 genes identified in the CRISPR screen with drug treatment by MAGecK analysis. n=3 independent biological cell cultures, p-values ≤ 0.05 after Benjamini and Hochberg multiplicity correction (FDR, one sided). **c.** Similarity between Tamoxifen and Fulvestrant CRISPR screens, identified significant genes which promote resistance or sensitivity.







Supplementary Figure 3

One way ANOVA test statistics for Extended data Fig. 1c (**a**) and e (**b**). n = 4 technical cell culture samples except MCF7 si*ARID1A* Veh n=3 and ZR-75-1 IBET762 = 5. Exact sample size is mentioned in the table.

a

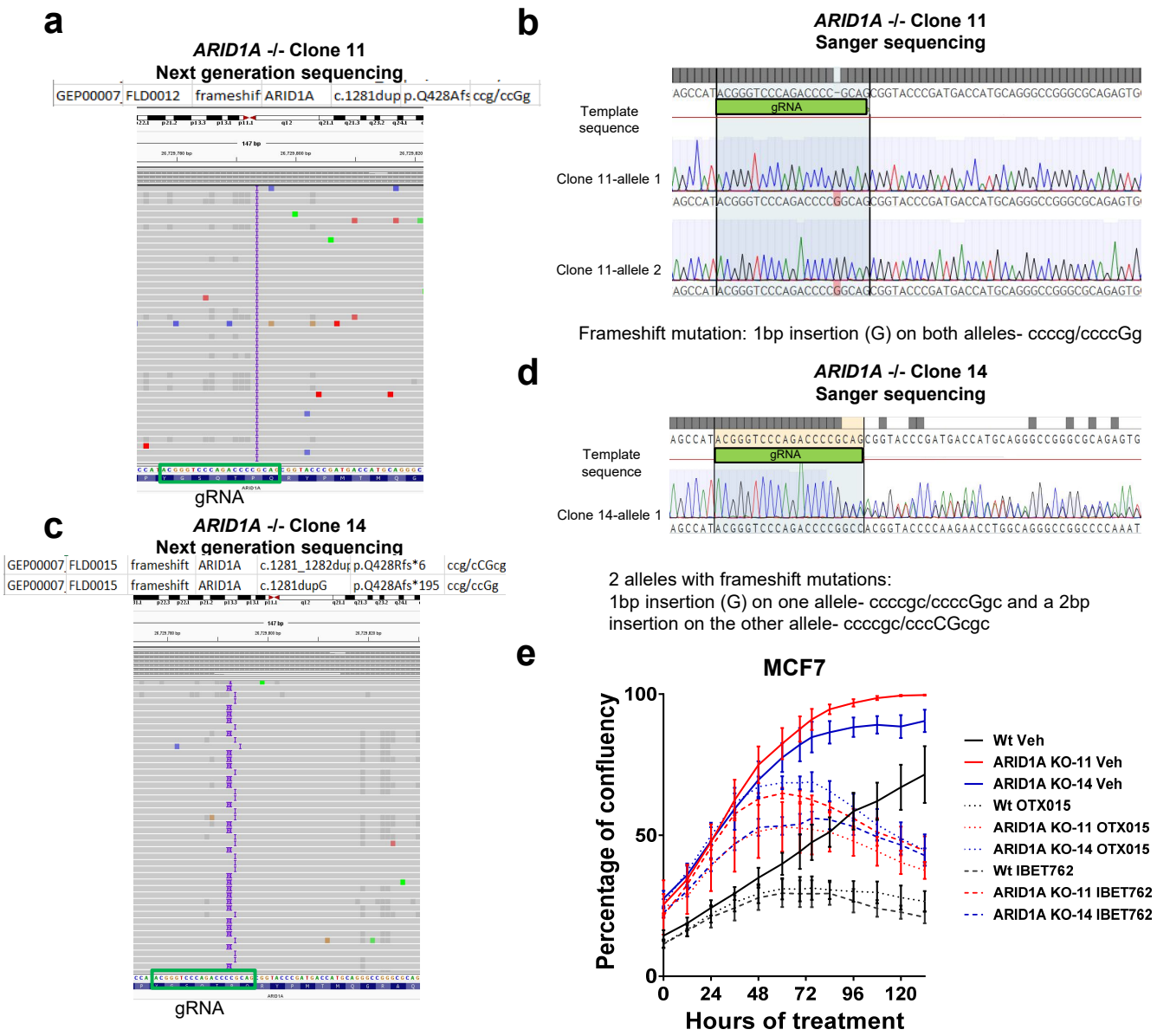
No.	Groups	Sample size	Source of Variation	SS	df	MS	F	P-value	F crit
1	siCont Veh	4	Between Groups	906.3526	1	906.3526	11.45514	0.019577	6.607891
	si*ARID1A* veh	3	Within Groups	395.6096	5	79.12191			
			Total	1301.962	6				
2	siCont Veh	4	Between Groups	2555.966	1	2555.966	67.34702	0.000177	5.987378
	siCont JQ1	4	Within Groups	227.713	6	37.95217			
			Total	2783.679	7				
3	siCont Veh	4	Between Groups	2532.439	1	2532.439	77.42859	0.00012	5.987378
	siCont ICI	4	Within Groups	196.2406	6	32.70677			
			Total	2728.68	7				
4	siCont ICI	4	Between Groups	1114.206	1	1114.206	30.1591	0.001527	5.987378
	si*ARID1A* ICI	4	Within Groups	221.6656	6	36.94426			
			Total	1335.871	7				
5	siCont	4	Between Groups	17.58714	1	17.58714	7.664173	0.03249	5.987378
	siCont E2 +Tam	4	Within Groups	13.76833	6	2.294722			
			Total	31.35547	7				
6	siCont E2 +Tam	4	Between Groups	396.2727	1	396.2727	13.45523	0.010477	5.987378
	si*ARID1A* E2 +Tam	4	Within Groups	176.7072	6	29.4512			
			Total	572.9799	7				

b

No.	Groups	Sample size	Source of Variation	SS	df	MS	F	P-value	F crit
1	siCont Veh	4	Between Groups	0.05032974	1	0.05032974	2.59986933	0.15799908	5.98737761
	si*ARID1A* Veh	4	Within Groups	0.11615139	6	0.01935857			
			Total	0.12692351	6				
2	siCont Veh	4	Between Groups	0.34899288	1	0.34899288	67.6002273	0.00017475	5.98737761
	siCont JQ1	4	Within Groups	0.0309756	6	0.0051626			
			Total	0.37996847	7				
3	siCont Veh	4	Between Groups	0.21081538	1	0.21081538	64.4544642	0.00019951	5.98737761
	siCont Tam	4	Within Groups	0.01962459	6	0.00327076			
			Total	0.23043997	7				
4	siCont Veh	4	Between Groups	0.50516863	1	0.50516863	129.033115	2.7884E-05	5.98737761
	siCont ICI	4	Within Groups	0.02349019	6	0.00391503			
			Total	0.52865881	7				
5	siCont Veh	4	Between Groups	0.24039441	1	0.24039441	59.5824422	0.00024799	5.98737761
	siCont OTX015	4	Within Groups	0.02420791	6	0.00403465			
			Total	0.26460233	7				
6	siCont Veh	4	Between Groups	0.03028669	1	0.03028669	7.43085425	0.0343698	5.98737761
	siCont IBET762	4	Within Groups	0.02445481	6	0.0040758			
			Total	0.0547415	7				
7	siCont JQ1	4	Between Groups	0.06030138	1	0.06030138	22.5258616	0.0031713	5.98737761
	si*ARID1A* JQ1	4	Within Groups	0.01606191	6	0.00267698			
			Total	0.07636328	7				
8	siCont Tam	4	Between Groups	0.5917724	1	0.5917724	178.779443	1.0832E-05	5.98737761
	si*ARID1A* Tam	4	Within Groups	0.01986042	6	0.00331007			
			Total	0.61163281	7				
9	siCont Fulv	4	Between Groups	1.68409741	1	1.68409741	459.127416	6.7406E-07	5.98737761
	siCont Fulv	4	Within Groups	0.02200824	6	0.00366804			
			Total	1.70610565	7				
10	siCont OTX015	4	Between Groups	0.23024385	1	0.23024385	122.42364	3.2444E-05	5.98737761
	si*ARID1A* OTX015	4	Within Groups	0.01128428	6	0.00188071			
			Total	0.24152814	7				
11	siCont IBET762	5	Between Groups	0.20829979	1	0.20829979	32.2072879	0.00030373	5.11735503
	si*ARID1A* IBET762	6	Within Groups	0.05820726	9	0.00646747			
			Total	0.26650705	10				

Supplementary Figure 4

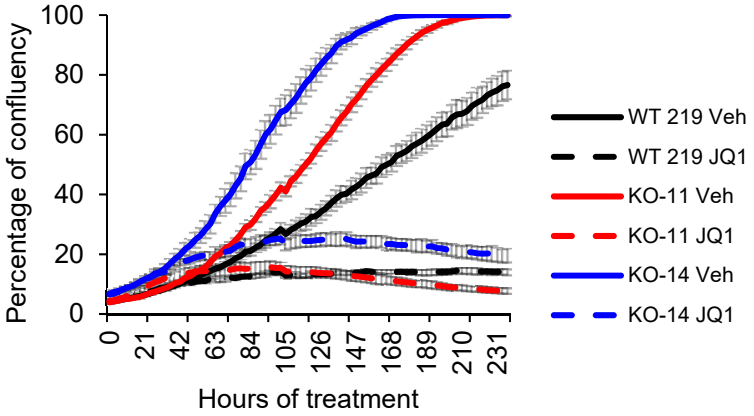
CRISPR knockout of *ARID1A* was confirmed by amplicon-based next generation sequencing (NGS) and Sanger sequencing for (**a, b**) clone 11 and (**c, d**) clone 14. **e**. Effect of BET inhibitors OTX015 and IBET762 on proliferation of *ARID1A* knockout clones. A representative data is shown from 2 independent experiments, n=4 technical cell cultures ± SEM.



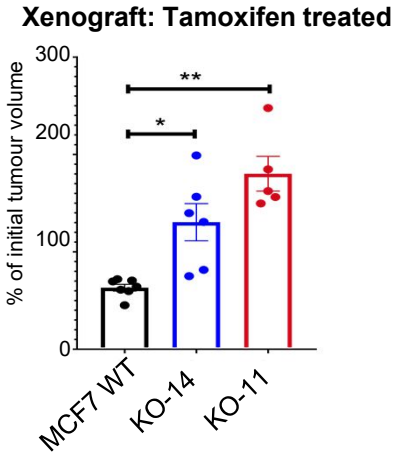
Supplementary Figure 5

a. Percent confluence as a function of hours of treatment in *in-vitro* proliferation assays using Incucyte in asynchronous MCF7 cells treated with vehicle or BETi (JQ1) for increasing time periods. Data of one representative experiment out of 4 is shown. n=4 ±SEM. **b, c.** Xenograft tumour growth of wild type MCF7 cells (n=7 animals) or *ARID1A* knockout clones (Clones 14 (n=6 animals) or Clone 11 (n=5 animals)) in tamoxifen-treated mice on day 25. p-values were calculated using Dunnett's T3 multiple comparisons test. * denotes p ≤ 0.05, ** denotes p ≤ 0.01. Bars correspond to the standard error of the mean. Test statistics are shown in **c**. **d.** Two sided-Wald Test and the statistics for Fig. 2c (details in Supplementary methods) shown for mice xenografts from MCF7 wild type (n=13 animals) and *ARID1A* knockout clones (Clones 11 (n=12 animals) or Clone 14 (n=8 animals). **e**. Images showing ARID1A Immunohistochemistry (IHC) on the xenografts derived from MCF7 *ARID1A* knockout cells Clone 11 and 14 from Fig. 2c and Fig. S5d. Clones were negative for ARID1A except the stromal cells from mice. All animals from Fig. S5d had been tested for ARID1A expression.

a



b



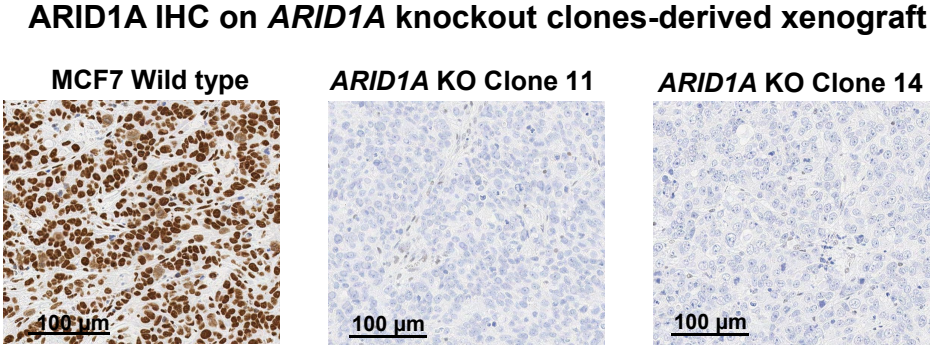
c

Dunnett's T3 multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value
MCF-7 TAM vs. KO-11 TAM	-117.0	-175.7 to -58.28	Yes	**	0.004
MCF-7 TAM vs. KO-14 TAM	-67.23	-125.0 to -9.438	Yes	*	0.03

d

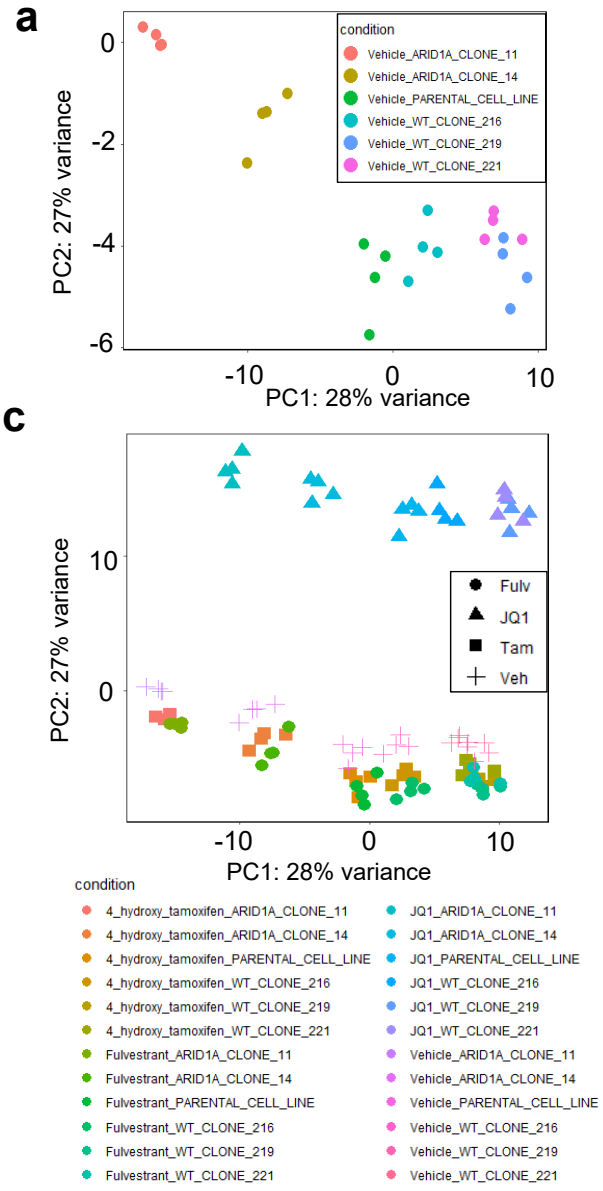
	Value	Std. Error	DF	t-value	p-value
(Intercept)	4.8104	0.2497	194	19.2630	$6.47X10^{-47}$
KO 11	0.1986	0.4346	30	0.4569	0.65105
KO 14	-0.0995	0.3789	30	-0.2625	0.79475
Days	0.1424	0.0129	194	11.0791	$2.03X10^{-22}$
KO 11 : Days	0.0964	0.0267	194	3.6096	0.00039
KO 14 : Days	0.0932	0.0213	194	4.3695	0.00002

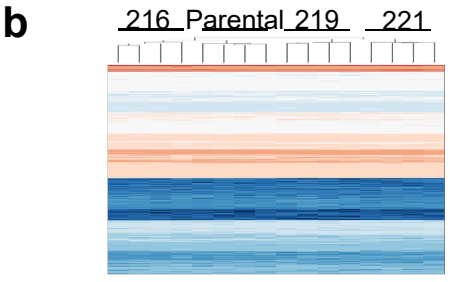
e

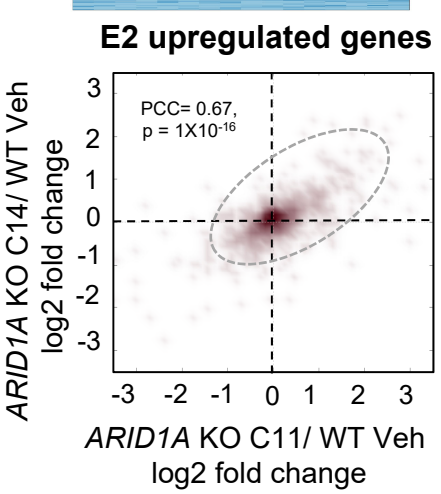


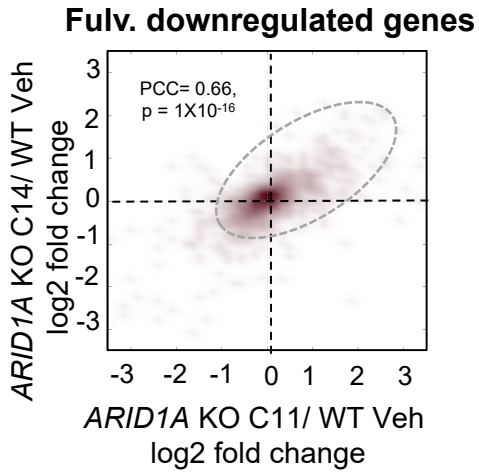
Supplementary Figure 6

Principal Component Analysis (PCA) (**a, c**) and hierarchical clustering (**b**) of gene expression with different clones of control and *ARID1A* knockout compared with parental cells and other treatments. n = 4 independent biological cell cultures. **d, e.** Scatterplot showing the gene expression in *ARID1A* knockout clones versus wild type control cells, specifically focusing on estrogen upregulated genes (**d**) or Fulvestrant downregulated genes (**e**) which represent direct ER target genes. PCC – Pearson Correlation coefficient, two sided. p-values were calculated by Pearson correlation test. **f**. Test statistics for the survival plot shown in Fig. 2E. Patients with up- (red) (n=104 for Vehicle and 72 for 4-hydroxytamoxifen) and down- (blue) (n=101 for Vehicle and 61 for 4-hydroxytamoxifen) regulation in gene expression.







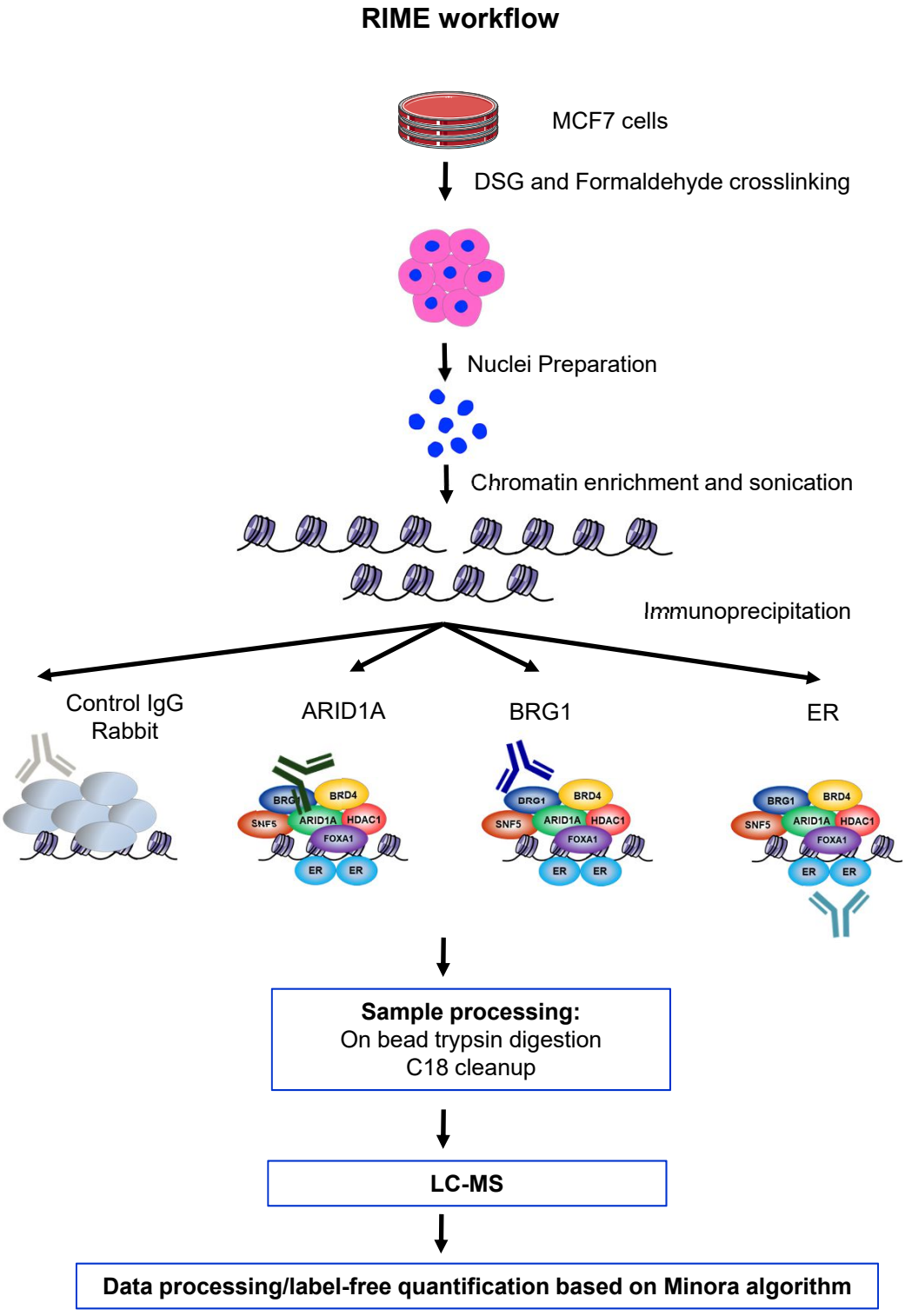


f

LOG-RANK test statistics for Fig 2e:

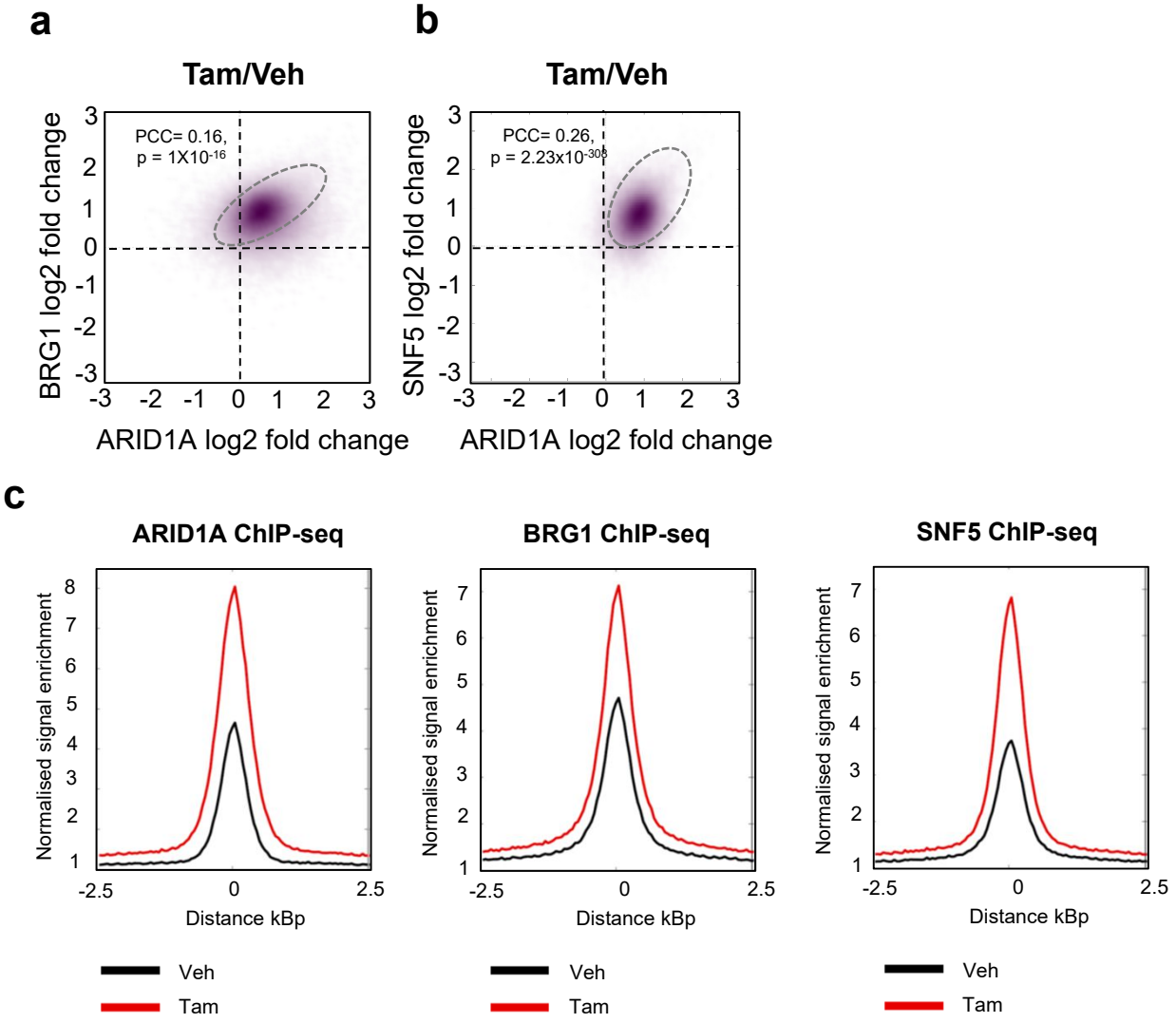
Treatment	UL	Standard error	Mantel-Haenszel Hazard ratio	95% confidence interval	UL normalized with Yates'es correction	two-sided p-value
Vehicle	12.16504	5.19907	1.5684	1.0758 - 2.2865	2.24368	0.02485
Tamoxifen	9.87179	4.43416	1.6522	1.0619 - 2.5705	2.11355	0.03455

Supplementary Figure 7



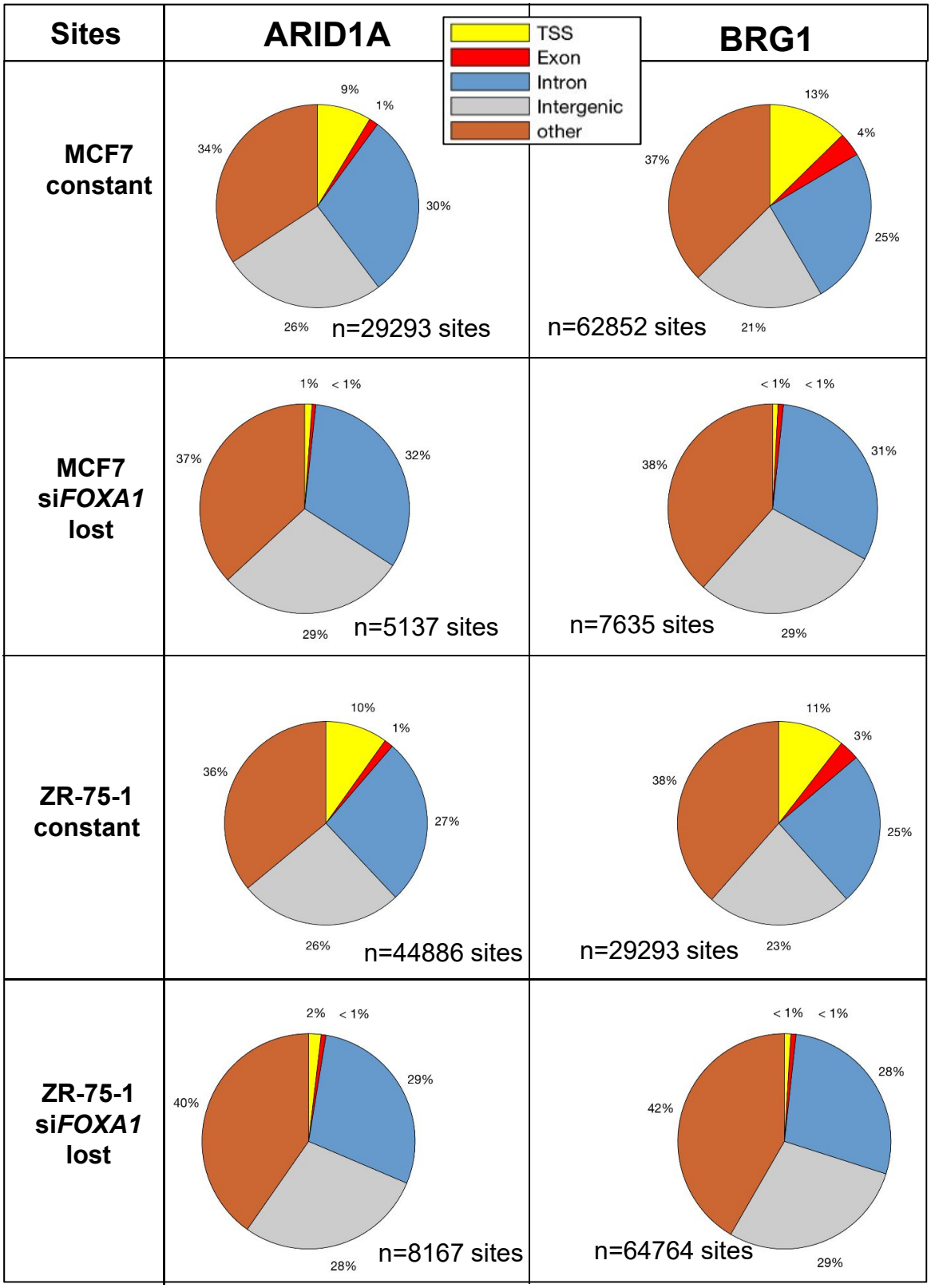
Supplementary Figure 8

a-c. Asynchronous MCF7 cells were treated with vehicle or Tamoxifen, and ChIP-seq was conducted for ARID1A, BRG1 or SNF5. 3 independent biological cell cultures were used. **a, b**. Scatterplot showing the association of ARID1A and BRG1 (**a**) or SNF5 (**b**) binding during Tamoxifen treatment. PCC – Pearson Correlation coefficient. p-values were calculated by Pearson correlation test, two-sided. **c.** Average density curves are shown on all ARID1A binding sites.



Supplementary Figure 9

Pie chart representing the percentage of TSS, exons, introns, intergeneic and other regions from the differential bound sites of ARID1A and BRG1 with si*FOXA1* in MCF7 and ZR-75-1 cells from 3 independent biological cell cultures. Number of sites regulated are mentioned in the figure.



Supplementary Figure 10

Test statistics for Extended Data Fig. 6c (**a**) and 6d (**b**). n = 3 independent biological cell culture samples.

a

One way ANOVA Comparison for Supplementary figure 15c

Cell line	Groups	Variance	Source of Variation	SS	df	MS	F	P-value	F crit
MCF7	ARID1A siCont	2.08552E-08	Between Groups	1.6E-06	1	1.6E-06	35.03457	0.004081	7.708647
	ARID1A si*FOXA1*	7.0241E-08	Within Groups	1.82E-07	4	4.55E-08			
			Total	1.78E-06	5				
MCF7	BRG1 siCont	5.24553E-08	Between Groups	3.91E-06	1	3.91E-06	112.0817	0.00045	7.708647
	BRG1 si*FOXA1*	1.72955E-08	Within Groups	1.4E-07	4	3.49E-08			
			Total	4.05E-06	5				
ZR-75-1	ARID1A siCont	4.34674E-07	Between Groups	8.59E-07	1	8.59E-07	3.788453	0.123467	7.708647
	ARID1A si*FOXA1*	1.88907E-08	Within Groups	9.07E-07	4	2.27E-07			
			Total	1.77E-06	5				
ZR-75-1	BRG1 siCont	3.20071E-08	Between Groups	1.89E-06	1	1.89E-06	28.01904	0.006115	7.708647
	BRG1 si*FOXA1*	1.03177E-07	Within Groups	2.7E-07	4	6.76E-08			
			Total	2.16E-06	5				

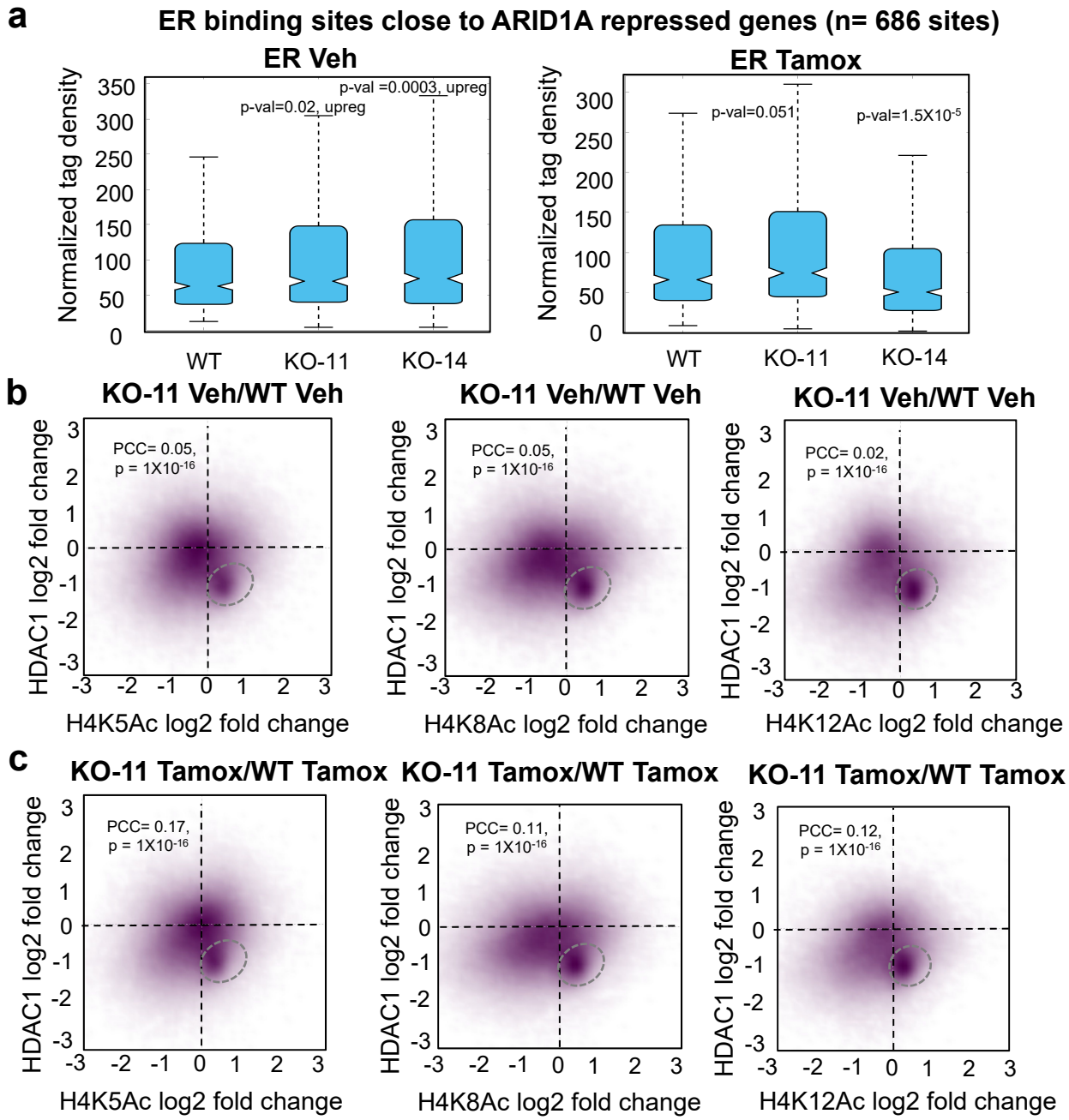
b

One way ANOVA Comparison for Supplementary figure 15d

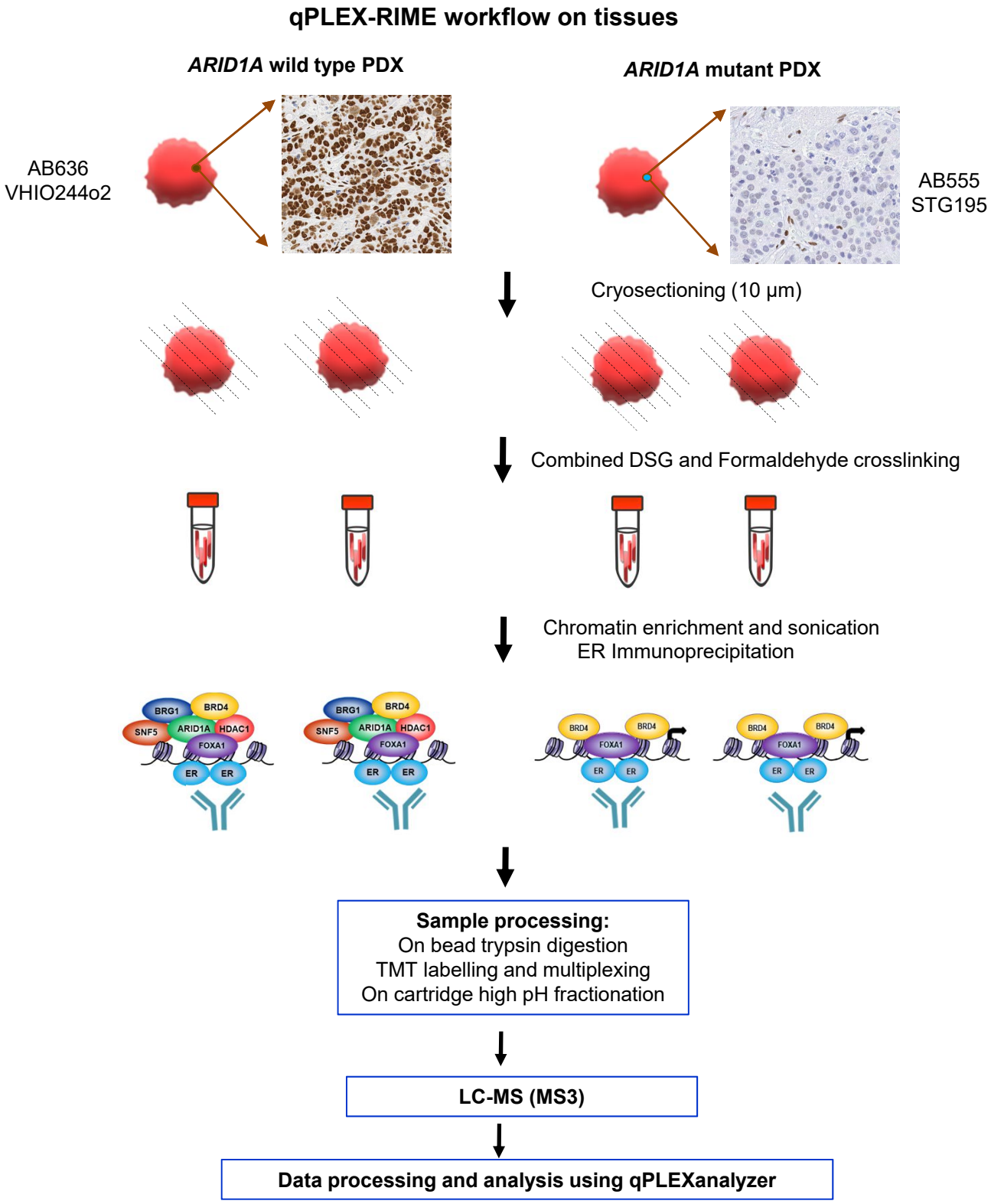
Site	Groups	Variance	Source of Variation	SS	df	MS	F	P-value	F crit
CCND1	siCont Veh	1.1E-05	Between Groups	6.85E-05	1	6.85E-05	9.804958	0.035142	7.708647
	si*FOXA1* Veh	2.92E-06	Within Groups	2.79E-05	4	6.98E-06			
			Total	9.64E-05	5				
CCND1	siCont Tam	4.25E-05	Between Groups	6.32E-05	1	6.32E-05	2.934406	0.161869	7.708647
	si*FOXA1* Tam	5.57E-07	Within Groups	8.62E-05	4	2.16E-05			
			Total	0.000149	5				
CDH1	siCont Veh	3.12E-06	Between Groups	7.7E-06	1	7.7E-06	4.522898	0.100579	7.708647
	si*FOXA1* Veh	2.86E-07	Within Groups	6.81E-06	4	1.7E-06			
			Total	1.45E-05	5				
CDH1	siCont Tam	6.81E-07	Between Groups	1.72E-05	1	1.72E-05	24.70636	0.007649	7.708647
	si*FOXA1* Tam	7.08E-07	Within Groups	2.78E-06	4	6.95E-07			
			Total	1.99E-05	5				

Supplementary Figure 11

(**a**) Boxplots of ER ChIP-seq signals showing no change in intensity (y-axis) with *ARID1A* knockouts clones 11 and 14. Plots were made on ER sites close to ARID1A repressed genes (n=686 sites) with more than 75% contribution to the variance in intensity. p-values were calculated by Welch test, two-sided. Window – 400 bp around ER binding. For boxplots, centre line shows the median values with bounds of box corresponding to the first and third quartiles and the upper and lower whiskers extend to the largest or the smallest value no further than 1.5 × IQR (inter-quartile range). More details are mentioned in Supplementary Table 5g. (**b, c**) Association of HDAC1 and histone H4 acetylation in wild type or *ARID1A* knock-out clone 11 in Vehicle (**b**) or Tamoxifen (**c**) treated cells. The binding is shown on all binding regions of HDAC1 and H4Ac marks, with log2 fold changes of knockout vs wild type for each treatment. HDAC1 log2 fold change was calculated only for Tamoxifen treatment. n =3 independent biological cell cultures. PCC- Pearson Correlation coefficient with p-values calculated by Pearson correlation test, two-sided.

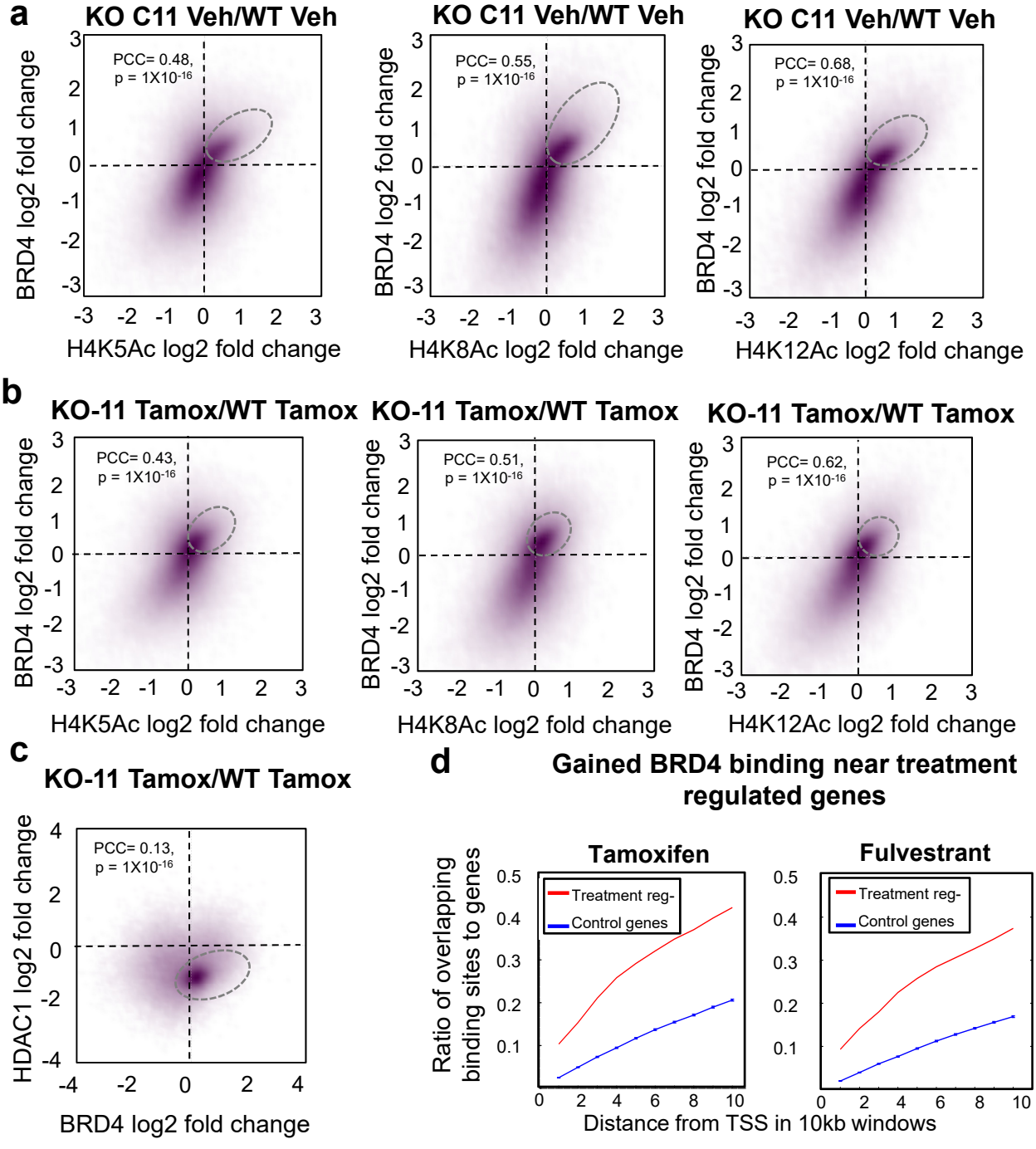


Supplementary Figure 12



Supplementary Figure 13

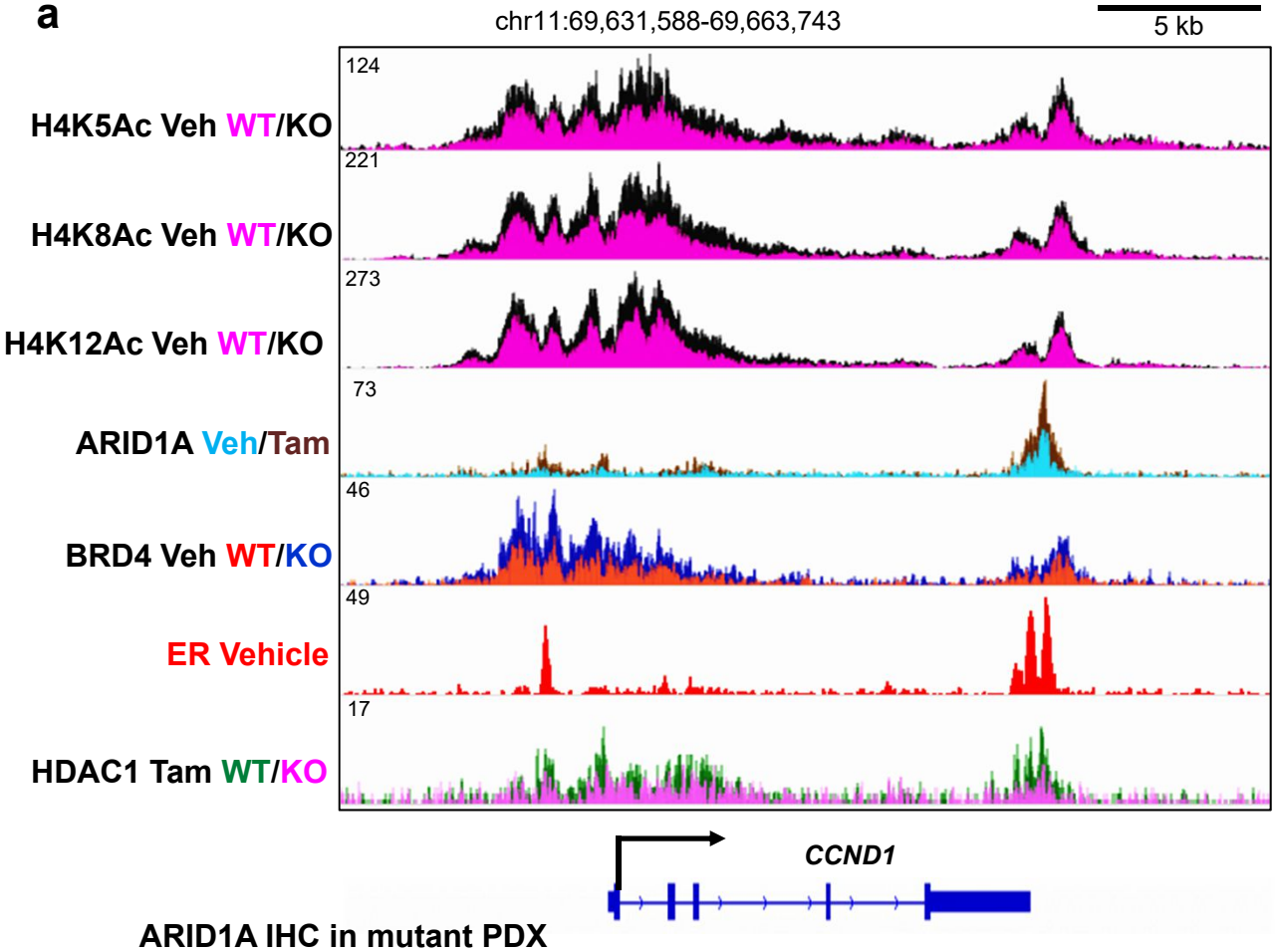
a, b. Association of BRD4 and histone H4 acetylation in *ARID1A* knock-out clone 11 with Vehicle (**a**) or Tamoxifen (**b**) comparing to wild type. The binding is shown on all binding regions of BRD4 and H4Ac marks with log2 fold changes of knockout vs wild type for each treatment. PCC- Pearson Correlation coefficient with p-values calculated by Pearson correlation test, two-sided. n =3 independent biological cell cultures. **c**. Association of BRD4 and HDAC1 in *ARID1A* knock-out clone 11 with Tamoxifen comparing to wild type. All binding regions of BRD4 and HDAC1 marks were shown with log2 fold changes of knockout vs wild type during Tamoxifen treatment. n =3 independent biological cell cultures. PCC- Pearson Correlation coefficient with p-values calculated by Pearson correlation test, two-sided. **d**. BRD4 gained binding events in *ARID1A* knock-out cells were positively associated with genes regulated by tamoxifen or Fulvestrant in an ARID1A-dependent manner. The curves represent ratio of binding sites in 10kb windows from the transcription start sites (TSS) of the genes.



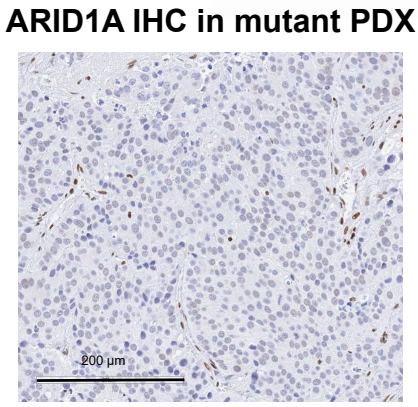
Supplementary Figure 14

a. Genomic profile showing ER, ARID1A, histone acetylation, BRD4 and HDAC1 binding at a single genomic locus, that encompasses *CCND1* gene, that are normally repressed by Fulvestrant/Tamoxifen, but not in the absence of ARID1A. **b**. ARID1A IHC staining in the *ARID1A* mutant PDX model used for explant study. IHC validated in 2 independent PDX passages and one explant study. **c**. Test statistics of the log rank test performed on Fig. 6d. ARID1A WT =93 patients, ARID1A mutant = 1731 patients.

a



b



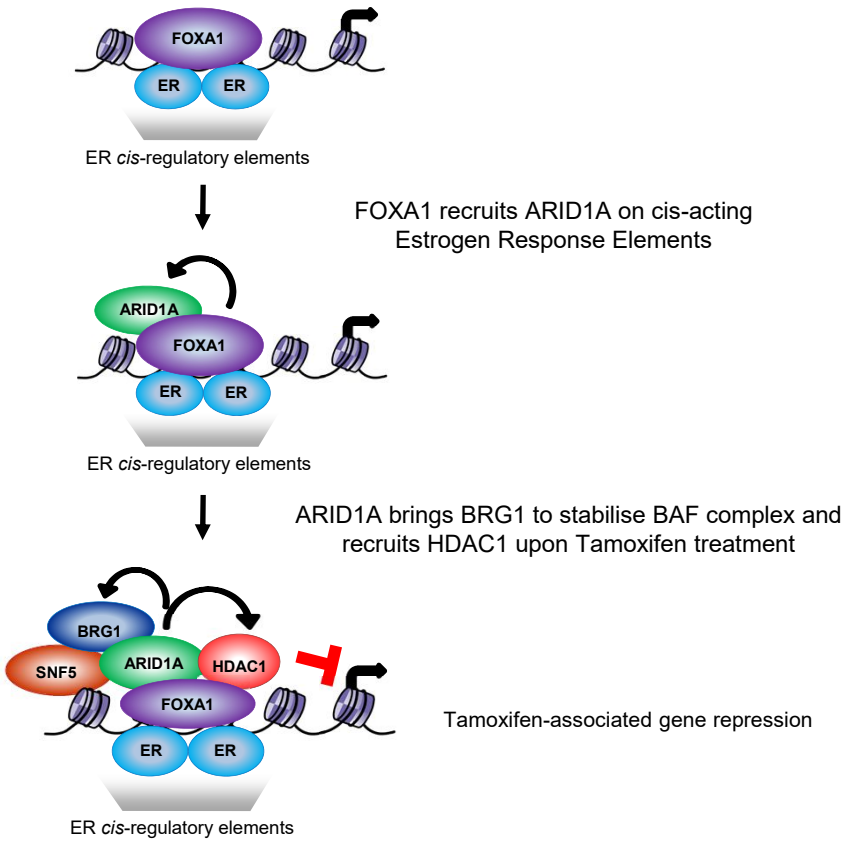
c

Treatment	UL	Standard error	Mantel-Haenszel Hazard ratio	95% confidence interval	UL normalized with Yates's correction	Two-sided p-value
ARID1A mutation	13.415	6.1933	1.4187	1.0338 - 1.9468	2.0853	0.037045

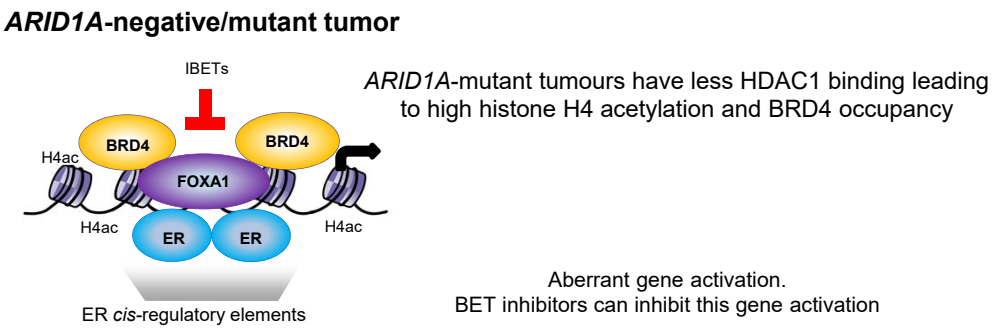
Supplementary Figure 15

Model describing the function of ARID1A in Tamoxifen-repressed transcriptional response (**a**) and its absence leads to aberrant gene activation (**b**).

a



b



Supplementary Note

Materials and Methods

Cell culture, treatments and transfection

MCF7 cells were obtained from European Collection of Authenticated Cell Cultures (ECACC) after testing for Mycoplasma contamination by RNA capture ELISA method (Mycoprobe™ from R&D systems) and profiled using Short Tandem repeats (STR) genotyping and Para DNA profiling. ZR-75-1 cells were obtained from American Type Culture Collection (ATCC). For the STR profiling, a commercially available 16 markers profile was utilized and analysed with Genemapper 5. Para DNA profile was obtained from LGC and the data were analysed with Para DNA analyser. Cells were tested for their response to ER antagonists and estrogen response regularly. All the cell lines and derived clones were grown in 10% FBS containing DMEM with high glucose, 2 mM Glutamine, Sodium Pyruvate, Penicillin and Streptomycin. ZR-75-1 cells were grown in RPMI media with the above supplements. Prior to estrogen treatment, cells were treated with 5% charcoal stripped serum (FBS) for 72 hours.

For RNA-seq and ChIP-seq experiments, treatments were done as follows: 100 nM (Z)-4-Hydroxy-Tamoxifen (Sigma H7904), 10 nM Fulvestrant (Selleckchem S1191), 250 nM JQ1 (Cayman chemical 11187), 10 nM β-estradiol (Sigma E2758) for 6 hrs. Proliferation experiments were performed in 96 well plates by seeding 2000-3000 cells, the treatments were started after 16 hrs of cell seeding and cells were grown for 7-10 days. 1 μM 4-Hydroxy-Tamoxifen was used. IBET762 (Selleckchem S7189) and OTX015 (Cayman chemical 15947) were treated as 2 μM and 500 nM respectively. Percentage of confluency was recorded every 3 hrs by Incucyte® Zoom Live cell analysis system from Essen Bioscience. Cell viability of ZR-75-1 were detected using Promega CellTiter-Glo Luminescent Cell Viability Assay as ZR-75-1 cells grow on top of each

other and confluency analyses is not a good fit. Non-targeting (D-001210-02) and *ARID1A* siRNAs (M17263-01-0005) targeting 5'-GCAACGACAUGAUUCCUAU-3', 5'-GAAUAGGGCCUGAGGGAAA-3', 5'-AGAUGUGGGUGGACCGUUA-3' and 5'-UAGUAUGGCUGGCAUGAUC-3' were obtained from Dharmacon smartpool siGenome siRNAs. FOXA1 siRNA targeting 5'-GCACUGCAAUACUCGCCUU-3', 5'-CCUCGGAGCAGCAGCAUAA-3', 5'-GAACAGCUACUACGCAGAC-3', 5'-CCUAAACACUUCCUAGCUC-3' and its corresponding control Non-targeting smartpool siRNA (D-001810-10) were used from Dharmacon smartpool ON-TARGETPlus siRNAs. Cells were transfected using Lipofectamine RNAimax (Thermo Scientific).

Western blotting

Whole cell lysates were lysed using Pierce RIPA buffer with complete EDTA-free protease inhibitor (Roche) and Halt Phosphatase inhibitor (Thermo Scientific), sonicated in Diagenode Bioruptor® Plus for 2-3 cycles in high power with 30 secs on/off. Protein was quantified by Millipore Direct detect® assay-free cards. SDS-PAGE was run using NuPAGE 4-12% gradient Bis-Tris gels, transferred using Bio-Rad wet transfer apparatus with methanol-containing NuPAGE transfer buffer in 100 V for 90 mins. Blocking was done using TBS-Odyssey blocking buffer (1:1) for 1 hr and primary antibodies were incubated overnight at 4°C. Antibodies used – ARID1A HPA005456 (Human Protein Atlas) or D2A8U 12354 (CST), β-actin ab6276 (abcam) and ER-α NCL-L-ER-6F11 (Leica Biosystems Novocastra). LI-COR CLx was used to visialise the fluorescent probed proteins.

ChIP-sequencing

ChIP was performed as previously published(*2, 3*). Double crosslinking was performed for all ChIPs except histone modifications. 2 X 15 cm dishes were crosslinked with 2 mM Disuccinimidyl glutarate (DSG from Santa Cruz sc-285455A) for 20 minutes after removing the media. After

removing DSG from cells, they were again crosslinked with 1% methanol free formaldehyde (Thermo Scientific) for 10 minutes. These were quenched by 100 mM Glycine, washed twice with ice-cold PBS and scraped to collect the cells. Cross-linked cells were washed with buffers containing complete EDTA-free protease inhibitor (Roche) and Halt Phosphatase inhibitor (Thermo Scientific): lysis Buffer-1 (50 mM Hepes–KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40/Igepal CA-630, 0.25% Triton X-100) and lysis Buffer-2 (10 mM Tris–HCL, pH8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) each for 10 minutes. Chromatin was suspended in lysis Buffer-3 (10 mM Tris–HCl, pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na–Deoxycholate, 0.5% N-lauroylsarcosine). This was sonicated for 15-20 cycles on high power 30 seconds on and off in a Bioruptor Plus (Diagenode). Sonicated chromatin was verified to have 200-600 bp fragments in an Agarose gel electrophoresis after reverse crosslinking by incubating 10 µl of chromatin at 95°C and purifying the DNA using Qiagen PCR purification kit. Chromatin was added with 1% Triton X-100 and centrifuged for 10 mins at 20,000g. These were aliquoted for IP after diluting with 1% Triton X-100 containing lysis buffer-3. 25 µl of input was taken and the remaining chromatin was added with 5 µg of antibody incubated with 50 µl dynabeads protein A (Thermo Scientific) which was blocked with 5 mg/ml BSA in PBS, washed and suspended in 1% Triton X-100 containing lysis buffer-3. Antibodies used: ARID1A (HPA005456), BRG1 (ab215998), SNF5 (Bethyl A301-087A), ERα (ab3575 and Merck Millipore 06-935 antibody mix), H3K27ac (Diagenode C15410196 Premium), H4K5ac (ab51997), H4K8ac (ab15823), H4K12ac (Diagenode C15410331-50), HDAC1 (Diagenode Premium C15410325) and ARID2 (Bethyl A302-229A). The antibody-coupled beads were incubated with chromatin overnight.

Beads were washed with ice-cold modified RIPA buffer (50mM HEPES pH 7.6, 1mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5M LiCl) six times in cold room, ice-cold TE buffer (10 mM Tris pH8.0 and 1 mM EDTA) once and diluted with elution buffer (50mM TrisHCl, pH8, 10mM EDTA,

1% SDS) before reverse crosslinking at 65°C overnight. IP samples including input were treated with RNase at 37°C for 30 mins and Proteinase K at 55°C for an hour. DNA was extracted using phenol chloroform and isoamyl alcohol (Sigma) and precipitated using ethanol with NaCl and glycogen. Pellets were washed with ice-cold 70% ethanol and dissolved in 10-15µl 10 mM Tris-HCl pH 8.0.

ChIP-qPCR analyses were performed on Bio-Rad CFX Connect Real-time PCR detection system. qPCRs were performed on ER and ARID1A binding sites using primers close to the genes as noted in the Supplementary Table 6. For making ChIP-seq libraries, samples were processed with Takara Bio ThruPLEX DNA seq kit with 96 dual indices for Illumina sequencing. Samples were size-selected using double sided 0.55X-0.65X and 0.25X using Beckman Coulter Agencourt Ampure XP beads. HiSeq 4000 was used to yield 30 million 50 bp single end reads per sample. 1% PhiX version3 viral genome spike-in was introduced during sequencing.

ARID1A Immunohistochemistry

Assessment of ARID1A protein expression in xenografts and cell line models was performed using immunohistochemistry as previously described (Khalique et al JPathClin Res 2018). Briefly 5um FFPE sections were incubated with the anti-ARID1A, rabbit monoclonal antibody at 1:1000 dilution, EPR13501 (Abcam, Cambridge, UK using the Dako-Autostainer Link 48 with the EnVision FLEX kit as per manufacturer's instructions (Agilent Technologies, Cheadle, Cheshire, UK). Isogenic HCT116 wild-type and ARID1A knockout cells were used as positive and negative controls respectively as described (Khalique et al JPathClin Res 2018). Stromal cell reactivity was used as an internal positive control.

Ki67 Immunohistochemistry

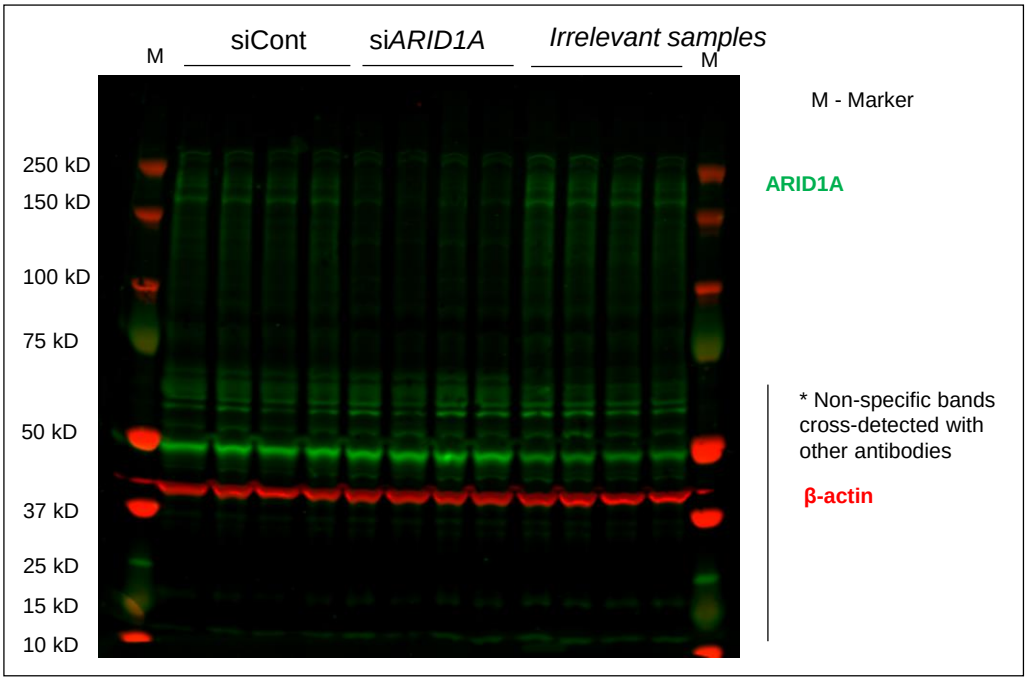
The de-waxing and re-hydration (as standard) prior to IHC as well as the post-IHC de-hydration and clearing were done on the automated Leica ST5020. Sections were mounted on Leica's coverslipper, CV5030. IHC was run using Leica's Polymer Refine Detection System (DS9800) using their standard template on the automated Bond-III platform. For xenograft samples, the MOM (mouse on mouse) protocol was used for anti-human ki67 to reduce background staining in the host tissue (because the antibody is a mouse monoclonal). This method includes an additional block (mouse Ig block solution, Vector MKB-2213) and an isotype specific secondary rabbit anti-mouse IgG1 (ab125913, diluted 1:1500) in place of the post primary antibody. Ki67 antibody was used from Dako, M7240 in the dilution of 1:400 with retrieval using Tris EDTA using 30 mins. The Tris EDTA pre-treatment is run at 100°C. DAB Enhancer is added as an ancillary reagent (Leica, AR9432).

Supplementary References:

1. K. Tzelepis *et al.*, A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep* **17**, 1193-1205 (2016).
2. D. Schmidt *et al.*, ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods*, (2009).
3. C. S. Ross-Innes *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
4. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21 (2013).
5. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
6. Y. Zhang *et al.*, Model-based Analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
7. T. L. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-208 (2009).
8. C. Curtis *et al.*, The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352 (2012).
9. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
10. H. Mohammed *et al.*, Progesterone receptor modulates ERalpha action in breast cancer. *Nature* **523**, 313-317 (2015).
11. M. M. Centenera *et al.*, Evidence for efficacy of new Hsp90 inhibitors revealed by ex vivo culture of human prostate tumors. *Clin Cancer Res* **18**, 3562-3570 (2012).
12. M. M. Centenera, G. V. Raj, K. E. Knudsen, W. D. Tilley, L. M. Butler, Ex vivo culture of human prostate tissue and drug development. *Nat Rev Urol* **10**, 483-487 (2013).

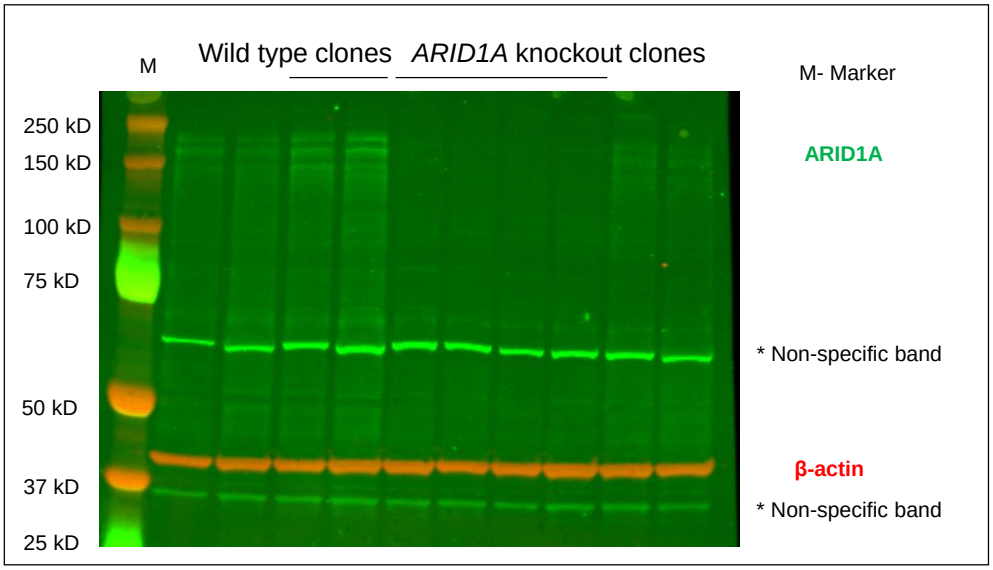
13. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308 (2013).
14. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754-1760 (2009).
15. Z. Lai *et al.*, VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* **44**, e108 (2016).

Western blot for Extended Data Fig. 1d



Western blot for Fig. 2a

ARID1A



ERα and β-actin