

## The pioneer factor PBX1 is a novel driver of metastatic progression in ER $\alpha$ -positive breast cancer

Luca Magnani<sup>1</sup>, Darren K. Patten<sup>1</sup>, Van T.M. Nguyen<sup>1</sup>, Sung-Pil Hong<sup>1</sup>, Jennifer H. Steel<sup>1</sup>, Naina Patel<sup>1</sup>, Ylenia Lombardo<sup>1</sup>, Monica Faronato<sup>1</sup>, Ana R. Gomes<sup>1</sup>, Laura Woodley<sup>1</sup>, Karen Page<sup>2</sup>, David Guttery<sup>2</sup>, Lindsay Primrose<sup>2</sup>, Daniel Fernandez Garcia<sup>2</sup>, Jacqui Shaw<sup>2</sup>, Patrizia Viola<sup>3</sup>, Andrew Green<sup>4</sup>, Christopher Nolan<sup>4</sup>, Ian O. Ellis<sup>4</sup>, Emad A. Rakha<sup>4</sup>, Sami Shousha<sup>1</sup>, Eric W.-F. Lam<sup>1</sup>, Balázs Győrffy<sup>5</sup>, Mathieu Lupien<sup>6,7</sup>, R. Charles Coombes<sup>1</sup>

<sup>1</sup>Department of Surgery and Cancer, Imperial College London, London, UK

<sup>2</sup>Department of Cancer Studies, University of Leicester, Leicester, UK

<sup>3</sup>Laboratory of Medicine, Histopathology Department, Royal Brompton Hospital, London, UK

<sup>4</sup>Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Budapest, HU

<sup>5</sup>MTA TTK Lendület Cancer Biomarker Research Group, 2nd Department of Pediatrics, Semmelweis University and MTA-SE Pediatrics and Nephrology Research Group, Budapest, HU

<sup>6</sup>The Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada

<sup>7</sup>Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

### Correspondence to:

Luca Magnani, e-mail: l.magnani@imperial.ac.uk

**Keywords:** breast cancer, estrogen receptor, drug resistance, metastasis, pioneer factors

**Received:** February 18, 2015

**Accepted:** June 02, 2015

**Published:** June 15, 2015

### ABSTRACT

Over 30% of ER $\alpha$  breast cancer patients develop relapses and progress to metastatic disease despite treatment with endocrine therapies. The pioneer factor PBX1 translates epigenetic cues and mediates estrogen induced ER $\alpha$  binding. Here we demonstrate that PBX1 plays a central role in regulating the ER $\alpha$  transcriptional response to epidermal growth factor (EGF) signaling. PBX1 regulates a subset of EGF-ER $\alpha$  genes highly expressed in aggressive breast tumours. Retrospective stratification of luminal patients using PBX1 protein levels in primary cancer further demonstrates that elevated PBX1 protein levels correlate with earlier metastatic progression. In agreement, PBX1 protein levels are significantly upregulated during metastatic progression in ER $\alpha$ -positive breast cancer patients. Finally we reveal that PBX1 upregulation in aggressive tumours is partly mediated by genomic amplification of the PBX1 locus. Correspondingly, ER $\alpha$ -positive breast cancer patients carrying PBX1 amplification are characterized by poor survival. Notably, we demonstrate that PBX1 amplification can be identified in tumor derived-circulating free DNA of ER $\alpha$ -positive metastatic patients. Metastatic patients with PBX1 amplification are also characterized by shorter relapse-free survival. Our data identifies PBX1 amplification as a functional hallmark of aggressive ER $\alpha$ -positive breast cancers. Mechanistically, PBX1 amplification impinges on several critical pathways associated with aggressive ER $\alpha$ -positive breast cancer.

### INTRODUCTION

Over 30% of ER $\alpha$ -positive breast cancer patients treated with endocrine therapies eventually relapse and progress into metastatic disease [1]. Previous studies

strongly suggest that the endocrine therapy resistance and metastatic progression occur in a multi-step manner and are driven by genetic and epigenetic alterations [2–4]. Several evidences indicate that breast cancer growth becomes ultimately independent from ER $\alpha$  [5–8].

Nonetheless this may be occurring toward the final stages of tumour progression. For instance, recent studies have discovered the existence of ER $\alpha$  activating mutations specifically in metastatic patients [9–12]. Additionally, ER $\alpha$  can be activated in the absence of estrogen by alternative mitogens including the epidermal growth factor (EGF) [13, 14]. EGF promotes ER $\alpha$  binding to thousands of regulatory elements and activates the transcription of a set of genes commonly over-expressed in the HER2 subtype and associated with disease progression [15].

We have recently described the role of PBX1 as a novel pioneer factor in ER $\alpha$  breast cancer [16]. PBX1 is critical to maintain chromatin accessibility and allow estrogen induced ER $\alpha$  recruitment to thousands of regulatory elements [16]. More importantly, we demonstrated that PBX1 is necessary for the transcriptional regulation of a subset of estrogen target genes directly associated with poor outcome. Furthermore, we have demonstrated that PBX1 upregulation is involved in the development of endocrine therapy resistance *in vitro* and that mRNA levels of genes controlled via Notch-PBX1 crosstalk are associated with poor prognosis in ER $\alpha$  patients [5]. Whether PBX1 contributes to estrogen-independent, EGF driven activation and recruitment of ER $\alpha$  is not yet known. We have thus investigated the role of PBX1 in mediating EGF signalling in ER $\alpha$ -positive breast cancer cell lines using genome wide-analyses including microarray and ChIP-seq. Our results demonstrate that PBX1 is required to direct EGF-ER $\alpha$  signalling at the chromatin level in breast cancer cells. PBX1 is also required for the expression of genes associated with tumour progression. PBX1 protein levels were significantly upregulated *in vivo* in metastatic samples from patients treated with endocrine therapies. Analysis of an independent historical collection of primary samples demonstrated that PBX1 protein levels correlate with shorter time to progression in ER $\alpha$  patients. Finally, we present evidences suggesting that PBX1 may be amplified in over 10% of ER $\alpha$ -positive breast cancer patients and amplification is associated with shorter survival. PBX1 amplification was confirmed in the circulating free DNA (cfDNA) of metastatic patients further suggesting the potential for using PBX1 to monitor breast cancer progression.

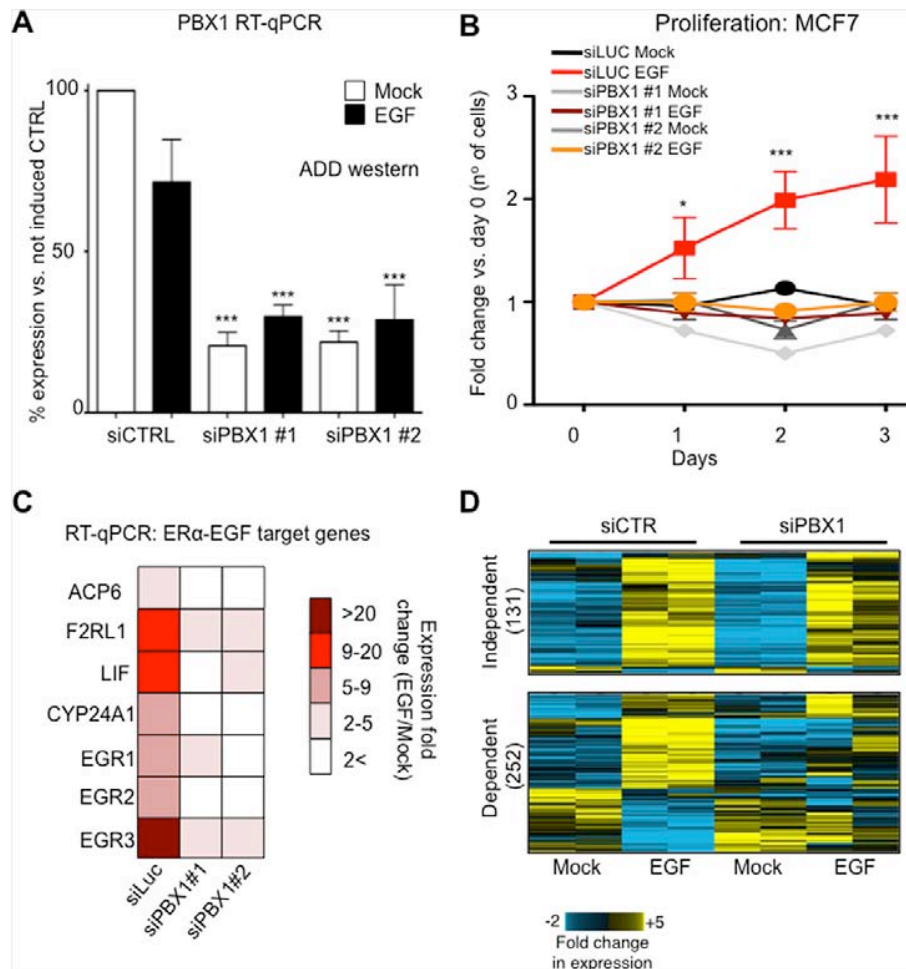
## RESULTS

### PBX1 mediates EGF signaling in ER $\alpha$ -positive breast cancer cells

We recently demonstrated that PBX1 is a pioneer factor in ER $\alpha$ -positive breast cancer, occupying the chromatin prior to ER $\alpha$  binding following its activation with estrogen and regulating the expression of estrogen dependent genes associated with aggressive progression

[16]. In addition we showed that ER $\alpha$  controls a different transcriptional program when activated by growth factors [15]. Specifically, EGF can activate ER $\alpha$  via serine phosphorylation and promotes transcription of genes associated with aggressive HER2 positive tumours [15]. Therefore, we reasoned that PBX1 might also play a role in estrogen independent EGF driven ER $\alpha$  induced transcription. ER $\alpha$ -positive MCF7 cells proliferate in response to EGF when cultured in the absence of estrogen and this phenotype is dependent on ER $\alpha$  [15]. PBX1 depletion using two siRNA molecules is sufficient to abrogate the proliferative response induced by EGF suggesting that PBX1 regulates EGF-ER $\alpha$  signaling (Fig. 1A–1B). Similar results were also obtained using T47-D, a second independent ER $\alpha$ -positive breast cancer cell line (Fig. S1A). We and others have previously showed that PBX1 is a direct Notch target that can be potentially antagonized with gamma secretase inhibitors (GSI, MRK003 and PF03084014) [5, 17]. In agreement, treatment with MRK003 is sufficient to downregulate PBX1 at mRNA and protein levels (Fig. S1B–S1C) and suppresses EGF induced proliferation in MCF7 cells (Fig. S1D–S1E). EGF activates ER $\alpha$  by promoting receptor phosphorylation at several key residues via AKT signaling pathway [15], including S118 [18]. Notably, PBX1 silencing did not impair ER $\alpha$  phosphorylation suggesting that defects in EGF induced proliferation did not stem from defect in non-canonical activation of ER $\alpha$  [19] (Fig. S2A). However, growth deficit in response to GSI treatment might stem from a combination of PBX1 downregulation and direct inhibition of EGF signalling (Fig. S2A).

Next, we investigated the transcriptional effect of PBX1 silencing on the EGF-ER $\alpha$  transcriptional program using qRT-PCR. As expected, silencing PBX1 in MCF7 cells impaired the transcriptional activation of a known set of EGF-ER $\alpha$  target genes [15] (Fig. 1C). These results were confirmed and expanded using expression microarrays (Fig. 1D). Our data indicate that over 60% (252 out of 383) of the genes differentially expressed in response to EGF signaling require PBX1 for their regulation (Fig. 1D and Table S1). These data are strongly reminiscent of our previous finding in which a subset of estrogen responsive genes were dependent on PBX1 [16]. PBX1 dependent genes were enriched for several important ontological terms associated with breast cancer and endocrine therapy resistance, including Notch signaling, pro-invasive signaling and epithelial and mammary carcinoma (Fig. 2). Interestingly, genes that failed to be induced by EGF in PBX1 depleted cells are also significantly enriched in several independent genes datasets obtained from patients characterized by aggressive breast cancers (Fig. S2B). Overall, these data suggest that PBX1 underlies the expression of EGF dependent genes involved in aggressive tumour progression.

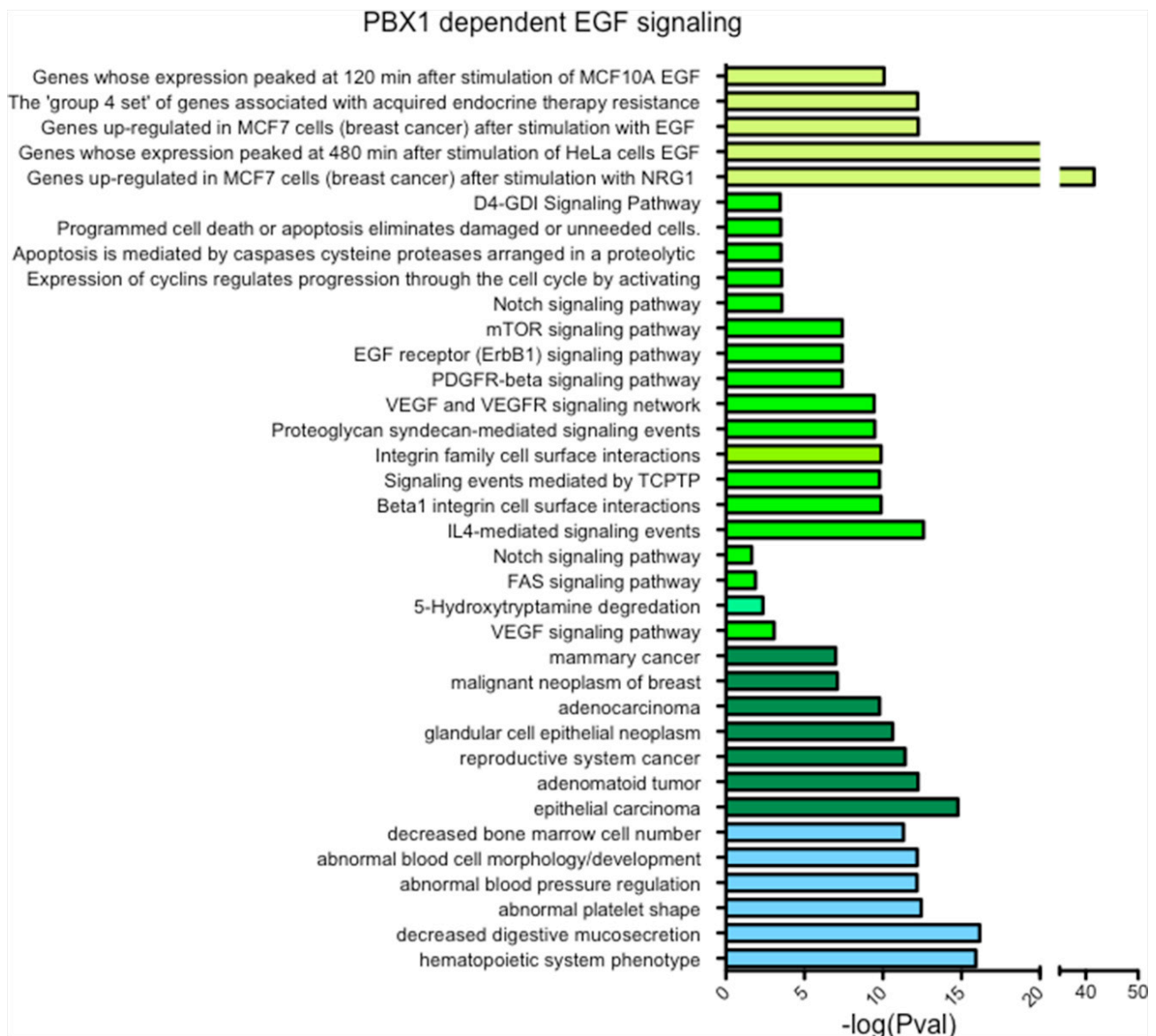


**Figure 1: PBX1 is required to activate EGF signaling in ER $\alpha$  breast cancer cells.** A. PBX1 silencing using two independent siRNAs B. Proliferation curves in response to EGF stimulation (100 ng/ml) or mock control in MCF7 cells. C. Transcriptional activation of several EGF target genes is lost upon PBX1 depletion. The fold induction was calculated averaging three independent experiments. All comparison between siRNA and control are statically significant (Student's *t*-test,  $p < 0.05$ ) D. Microarray analysis of PBX1 depleted cells demonstrates that PBX1 control a large portion of EGF dependent genes. Asterisks identify significant differences between treatments.

### PBX1 is pre-loaded at regulatory elements recruiting ER $\alpha$ during EGF signaling

Pioneer factors such as PBX1 and FOXA1 bind to the chromatin prior to ER $\alpha$  recruitment where they are thought to mediate or maintain chromatin accessibility [20–22]. To determine if PBX1 could serve as a pioneer factor for ER $\alpha$  under EGF stimulation we conducted ChIP-seq assays in MCF7 cells. We then compared the EGF-ER $\alpha$  cistrome (e.g. ER $\alpha$  genome wide binding sites induced upon EGF stimulation) with the estrogen-ER $\alpha$  cistrome and previously published PBX1 cistrome. In agreement with previous reports [15], most ER $\alpha$  chromatin binding is unique to specific stimuli (EGF-ER $\alpha$  vs. E2-ER $\alpha$ ) (Fig. 3A and S3A). PBX1 occupies approximately 30% of ER $\alpha$  binding sites, regardless of the stimulus (Fig. 3A). PBX1 binding occurs prior and is not affected by EGF stimulation as demonstrated by direct ChIP-qPCR assays

at selected binding sites in proximity of EGF regulated genes (Fig. 3B). This process is strongly reminiscent of PBX1 behaviour in the context of estrogen stimulation [16]. We next considered if EGF induced ER $\alpha$  binding sites that overlap with PBX1 have a functional relationship with the EGF-ER $\alpha$  transcriptional program. To do this we have combined transcriptional and ChIP-seq data of MCF-7 cells treated with EGF. We have analysed genes whose expression changed in response to EGF stimulation (all upregulated genes) for enriched binding patterns near their transcriptional starting site (up to 20 kb upstream of the TSS). We focused on ER $\alpha$  binding in response to estrogen (17 beta-estradiol, E2) or EGF with or without concurrent PBX1 and FOXA1 binding. Our data show that genes that respond to EGF are significantly enriched for PBX1-EGF-ER $\alpha$  binding but lack PBX1-E2-ER $\alpha$  binding sites (Fig. S3B, red dots). FOXA1 binding was less enriched near these genes (Fig. S3B, light-blue dots).



**Figure 2: PBX1 controls the expression of genes associate with breast cancer progression.** Genes whose failed to be upregulated by EGF upon PBX1 silencing ( $n = 147$ ) were used for ontology analysis using GREAT [48].

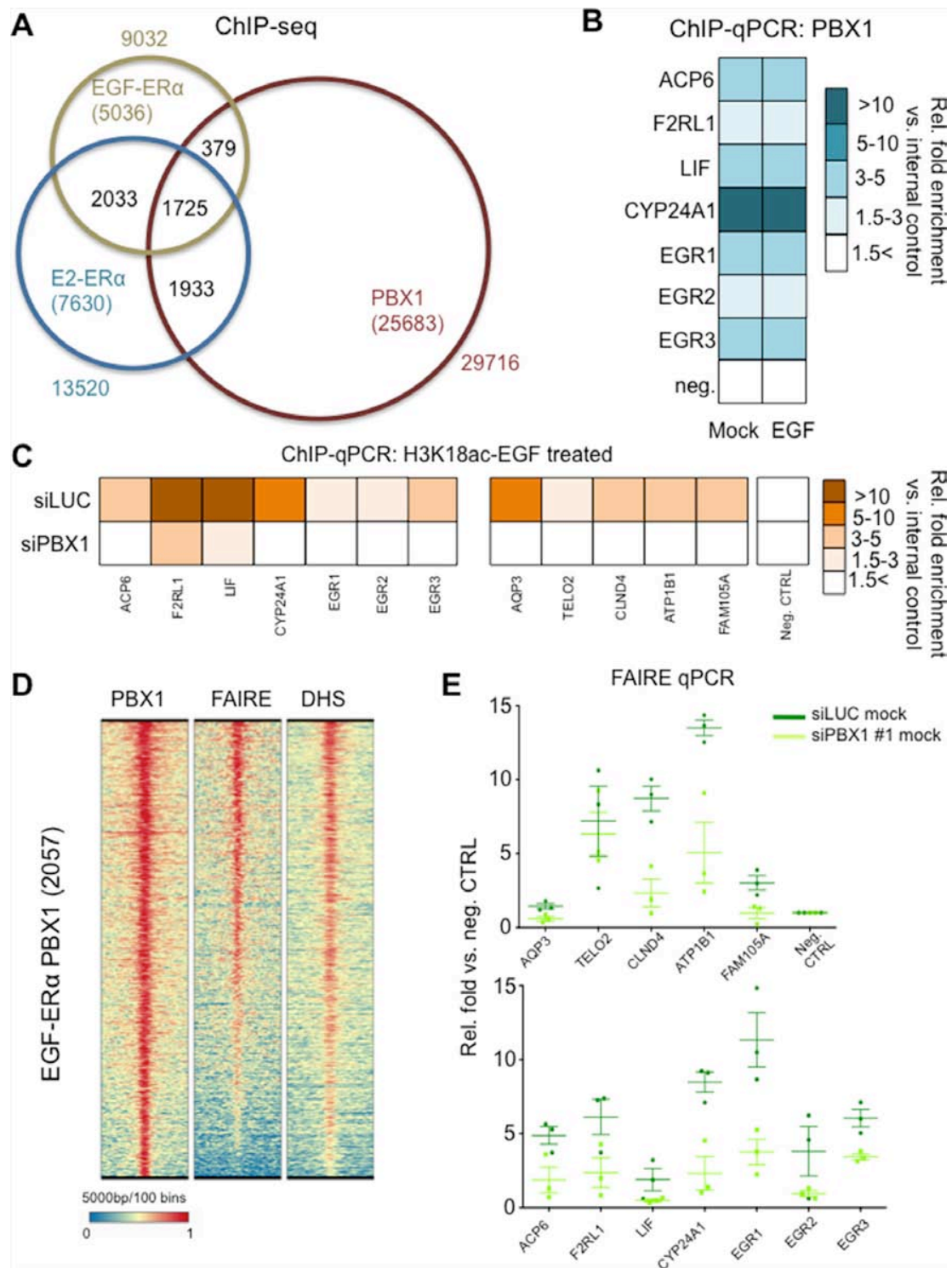
EGF responsive genes are also significantly enriched for PBX1-FOXA1-non specific ER $\alpha$  binding, suggesting that co-activation may also play an important role to regulate this subset of genes (Fig. S3, purple dots). As expected, regions pre-bound by pioneer factors that did not recruit ER $\alpha$  lacked any association with transcription. Collectively, these analyses confirm that PBX1 binding is directly involved in regulating EGF-ER $\alpha$  signaling in MCF-7 cells.

We have previously demonstrated that H3K18ac can be used to detect stimuli specific co-activation (EGF vs. E2) at regulatory regions in which ER $\alpha$  binding occurs under both treatment [15]. Interestingly PBX1 depletion directly suppressed H3K18ac signal at several putative elements normally recruiting ER $\alpha$  (Fig. 3C). Globally,

PBX1-EGF-ER $\alpha$  regions are characterized by open chromatin status as demonstrated by FAIRE-seq and DNaseI-hypersensitivity-seq assays (Fig. 3D). However, PBX1 depletion results in a significant decrease in chromatin openness (as demonstrated by reduced FAIRE signal) at several regions associated with EGF-ER $\alpha$  target genes suggesting that PBX1 modulates ER $\alpha$  signaling by influencing the chromatin environment.

### **PBX1 is a novel luminal breast cancer prognostic biomarker**

Our mechanistic study revealed that PBX1 controls the expression of a subset of ER $\alpha$  target genes stimulated



**Figure 3: PBX1 regulate chromatin status at EGF-ER $\alpha$  binding sites.** **A.** ChIP-seq analysis of PBX1 and ER $\alpha$  in MCF7 cells **B.** ChIP-qPCR validation of PBX1 binding at potential regulatory EGF-ER $\alpha$  binding sites in the absence of EGF stimulation. CTRL indicate stimulation using mock. The fold enrichment was calculated averaging three independent experiments. All comparison between siRNA and control are not statically significant (Student's *t*-test,  $p < 0.050$ ) **C.** PBX1 silencing abrogates epigenetic co-activation of EGF-ER $\alpha$  regulatory elements as measured by induced H3K18 acetylation. The fold enrichment was calculated averaging three independent experiments. All comparison between siRNA and control are statically significant (Student's *t*-test,  $p < 0.050$ ) **D.** Genome wide chromatin accessibility analyses at PBX1 and EGF-ER $\alpha$  shared sites. **E.** PBX1 depletion negatively impact chromatin accessibility at EGF-ER $\alpha$  binding sites as identified by reduced FAIRE signal.

by EGF *in vitro* and may be linked to breast cancer progression. Furthermore we previously demonstrated that PBX1 displays an analogous function by controlling a subset of estrogen induced genes strongly associated with poor prognosis [16]. In light of these observations we hypothesized that PBX1 expression levels could contribute to breast cancer progression in ER $\alpha$ -positive breast cancer patients. To test this hypothesis we performed meta-analysis on METABRIC (Illumina arrays) and The Cancer Genome Atlas (TCGA, RNA-seq data) studies stratifying ER $\alpha$ -positive breast cancer patients in high and low PBX1 expressors. In both datasets high PBX1 mRNA levels correlate with shorter survival (Fig. S4A). For the METABRIC dataset, PBX1 expression retained significance in a multivariate analysis including lymph node status, grade and tumor size (HR = 1.27,  $p$  = 0.0173). For the TCGA datasets the TNM status was published for each patient. In the multivariate analysis including PBX1, T, N and M, PBX1 reached the highest significance ( $p$  = 0.041, HR = 7.82), M also reached significance ( $p$  = 0.045, HR = 2.89) while T and N were not significant. These correlations were not significant in ER $\alpha$ -negative breast cancer patients further supporting the notion that PBX1 is a key regulator of ER $\alpha$  activity (Fig. S4A). Notably, higher PBX1 levels were significantly associated with poor prognosis in a subset of patients treated with endocrine therapy (Fig. S4B). Again, PBX1 retained significance in a multivariate analysis including lymph node status, grade, tumor size and PBX1 expression (HR = 1.27,  $p$  = 0.0173). To elucidate further the potential role of PBX1 as a breast cancer biomarker, we tested the association between its expression and metastatic progression in an additional six independent studies (Affymetrix arrays). Overall, PBX1 expression was associated with increased risk in metastatic progression (HR 1.47,  $p$  < 0.008,  $n$  = 674). We found more variability within the individual cohorts, however, statistical significance was reached only for datasets in which PBX1 represents a poor prognostic marker (Fig. S5). Additionally, multivariate analysis in the grouped cohort indicates that PBX1 prognostic power is independent from Ki67 ( $p$  < 0.0001).

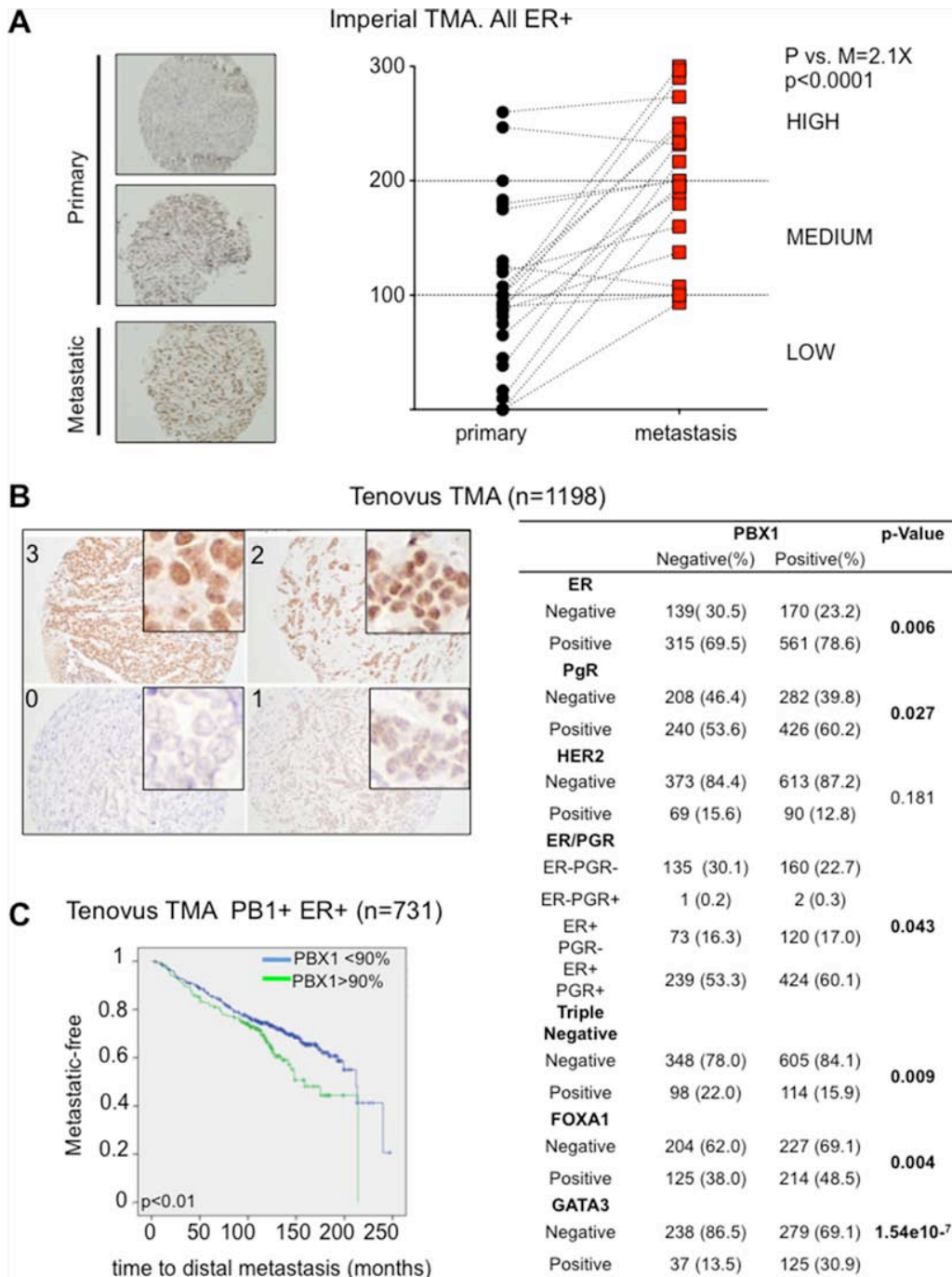
To confirm these findings we measured PBX1 protein levels using IHC in longitudinal clinical samples. We collected 20 biopsies from primary ER $\alpha$ -positive breast cancers and compared PBX1 protein levels with matched relapses (all endocrine treated, average time to relapse = 44.4  $\pm$  35 months). We stained all sections using a validated antibody (Fig. S1C and S2A) [5, 16] and scored each sample based on PBX1 nuclear intensity (Fig. 4A). Remarkably, metastatic samples had a significant increase in H score compared to the original primary biopsies confirming PBX1 upregulation in drug-resistant metastatic patients (Fig. 4A). Finally, we investigated PBX1 protein levels in the Tenovus Nottingham cohort of FFPE primary breast cancers ( $n$  = 1650) [23]. In agreement with cell

line data, we found that PBX1 protein levels significantly correlate with ER $\alpha$ , FOXA1 and GATA3 (another important ER $\alpha$  pioneer factor [22]) thus demonstrating that PBX1 is strongly associated with ER $\alpha$ -positive, luminal breast cancer subtypes (Fig. 4B). We therefore restricted further analysis to ER $\alpha$ -positive breast cancer patients. Since PBX1 protein levels were dramatically increased in metastatic samples, we hypothesized that high levels of PBX1 in primary samples might identify patient at higher risk of relapse. Our results show that ER $\alpha$ -positive breast cancer patients with elevated PBX1 staining intensity at diagnosis (over 90% positive staining) develop metastatic disease significantly faster than patients with lower staining intensity (Fig. 4C). Overall this data strongly suggest that PBX1 protein levels in primary breast cancer are significantly associated with breast cancer progression.

### **PBX1 locus is amplified in patients with aggressive ER $\alpha$ breast cancer**

Our analysis using PBX1 transcriptional and protein levels strongly supports the role of PBX1 as a functional biomarker associated with ER $\alpha$  breast cancer progression (Figs. 4 and S4–S5). PBX1 transcripts are significantly overexpressed in breast cancer when compared to normal tissues (Fig. S6). For example, PBX1 ranks in the top 50 most overexpressed genes in ductal and lobular carcinoma compared to normal tissue in TCGA cohort (dataset 30 in Fig. S6) [24]. Similarly, PBX1 ranks in the top 4% of the most overexpressed transcripts in the METABRIC cohort (Invasive Ductal vs. Normal, dataset 7, Fig. S6) [25]. To the best of our knowledge, there is no indication in the literature that PBX1 is required for normal breast development. In agreement, we found that normal ductal cells adjacent to tumor tissue have very low PBX1 staining (stronger stained cells are myoepithelial cells, Fig. S7). Thus, PBX1 expression seems to be a major determinant for the development and progression of luminal breast cancer patients.

Recent tumour sequencing efforts indicate that PBX1 locus is potentially amplified in 13% of primary breast cancers (TCGA provisional) (Fig. S8A) [26] supporting the hypothesis that genomic amplification may partially explain PBX1 overexpression in breast cancer (TCGA provisional) (Fig. S8B). Of note, within the published TCGA dataset [24], PBX1 potential amplification strongly correlates with higher mRNA levels in ER $\alpha$ -positive but not in ER $\alpha$ -negative breast cancer patients (Fig. S9). These data suggest that PBX1 copy number variation (CNV) might then be functional only in ER $\alpha$ -positive breast cancer patients. We then identified the TCGA patients (TCGA provisional) with ER $\alpha$ -positive disease that had at least 3 copies of PBX1 (95/1145, 8.3%) and analysed the proportion of patients with lymph-node micro-metastasis. Strikingly, we found that 60% of patients with PBX1 amplification had at



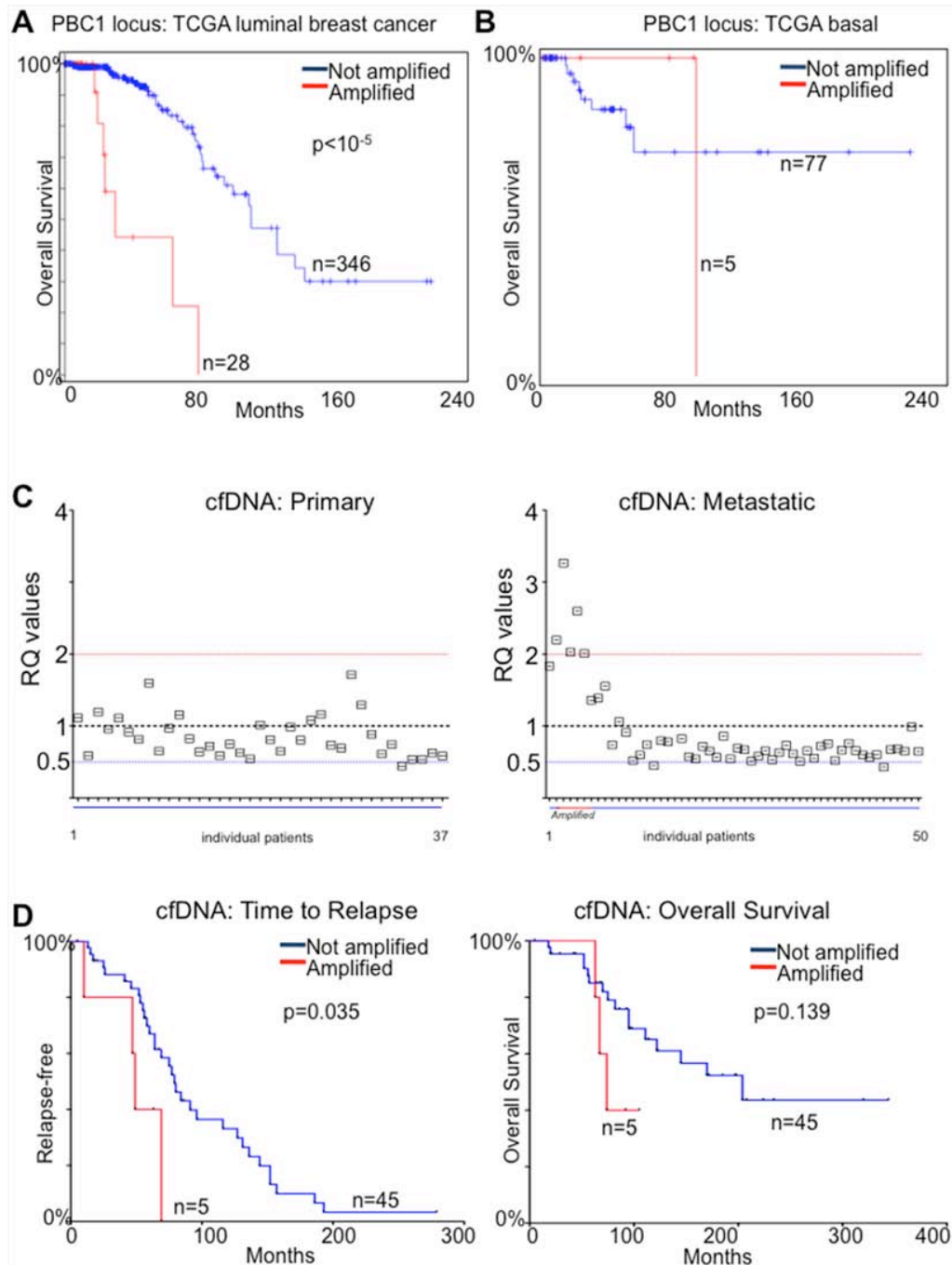
**Figure 4: PBX1 is a novel prognostic biomarker for ER $\alpha$  breast cancer.** **A.** Primary and matched Metastatic samples from the Imperial College TMA were processed using PBX1 IHC and H score were plotted. Pair-wise *t*-test between the average score (3 independent scorer, duplicate sections) was used to establish significance **B.** Core primary biopsies in the Tenovus Nottingham cohort show different level of staining by PBX1 IHC. PBX1 correlates with several important clinico-pathological parameters including hormone status and ER $\alpha$  associated pioneer factors. **C.** Survival analysis (Log-rank Mantel-Cox test) was performed looking at patients with very high PBX1 levels versus the remaining patients. Analysis was restricted to the to the ER $\alpha$ -positive population.

least one positive lymphnodes vs. 38.7% of the control group (Fisher exact test  $7.3 \times 10^{-5}$ ). Conversely, PBX1 amplification correlates, although not significantly, with

a lower chance of metastatic disease in the lymphnodes in ER $\alpha$ -negative breast cancer patients (45% vs. 57% respectively).

We then restricted our analysis to ER $\alpha$ -positive breast cancer patients from the published TCGA data to exploit the more complete clinical annotations [24] and found that 28/364 (8%) carried potential PBX1 amplification/overexpression (Fig. 5A). Patients with

putative PBX1 amplification had a median survival of 30.98 months compared to 113.74 (PBX1 Amplified vs. Not Amplified in Luminal patients,  $P < 10^{-5}$ ) (5A). On the other hand, PBX1 amplification did not reach prognostic significance in ER $\alpha$ -negative patients (however, only



**Figure 5: PBX1 amplification is a prognostic biomarker for ER $\alpha$  breast cancer.** A. TCGA luminal patients [24] were stratified based on putative PBX1 amplification and overall survival was calculated B. The same analysis was repeated in basal and claudin low patients C. qPCR measurement of PBX1 genomic DNA in cfDNA from primary and metastatic ER $\alpha$  breast cancer patients E–F. Metastatic patients from panel C were stratified based on PBX1 amplification and Time to relapse and Overall survival were calculated. Groups were compared using a Log-rank Mantel-Cox test.



6% of patients carries amplification) (Fig. 5B). When we looked at genes potentially co-amplified with PBX1 (chr1q23.3) [27] we found that very few candidates were also over-expressed in breast cancer (Fig. S10A). More importantly, PBX1 was one of the only two transcripts (the other being PVRL4) carrying evidence of prognostic significance in the ER $\alpha$ -positive breast cancer subtype (METABRIC) (Fig. S10B).

We then considered the possibility of using PBX1 CNV as a proxy for PBX1 protein levels. Potentially, this would allow for monitoring PBX1 levels by measuring its amplification in DNA derived from cancer cells using non-invasive methods. The tumour derived fraction of circulating free-DNA (cfDNA) originating from tumour cells can be used to estimate the genomic status of breast cancer lesions [28–30]. We then measured PBX1 amplification using a qPCR assay in cfDNA from 37 patients with ER $\alpha$ -positive primary disease and 50 patients with ER $\alpha$ -positive metastatic disease treated with endocrine therapy (Table S2). We could not identify significant CNV in blood from primaries ER $\alpha$ -positive breast cancer patients. However, we identified 5/50 (10%) patients with PBX1 amplification (RQ > 2) and two patients with probable allelic loss (4%) in the metastatic cohort (Fig. 5C). We then reasoned that the emergence of PBX1 CNV may have contributed to tumor progression and used PBX1 amplification in cfDNA as a classifier to stratify patients based on time to relapse or overall survival. In this small cohort, PBX1 amplification was significantly associated with earlier metastatic relapse ( $p < 0.035$ , HR 5.95, Cox- Mantel-Test) (Fig. 5D). These data are in strong agreement with our protein and transcriptional tumour profiling and coupled with our mechanistic study demonstrate that PBX1 plays a crucial role in ER $\alpha$ -positive breast cancer development and progression.

## DISCUSSION

Our work has identified PBX1 as a novel functional biomarker in ER $\alpha$ -positive breast cancer. We have identified a mechanistic association between PBX1 and ER $\alpha$  recruitment in response to EGF signaling. We also established the significance of PBX1 mRNA and protein level as a prognostic biomarker in several independent large cohorts of primary breast cancers. Notably, PBX1 levels are strongly associated with metastatic progression. Finally, we demonstrate that the longitudinal increment of PBX1 levels observed throughout cancer progression are partly due to genomic events that could be potentially used to monitor disease progression in patients by non-invasive assays.

The number of signaling molecules potentially involved in breast cancer is rapidly increasing. In this context, ER $\alpha$  remains a central hub required to regulate and integrate signaling at the chromatin level [31, 32].

EGF signaling cascade partially converge on ER $\alpha$  in breast cancer cells where it provides for an additional survival pathways. ER $\alpha$  target genes induced via classical estrogen signaling and non-canonical EGF signalling are quite distinct [14, 15]. Nonetheless, PBX1 remains a common denominator of these two pathways via its conserved interaction with ER $\alpha$ . It will be interesting to investigate the role of pioneer factors in alternative ER $\alpha$  pathways, including metabolic induced ER $\alpha$  signaling [33, 34]

Genetic aberration targeting chromatin remodelers and other proteins involved in epigenetic processes are now frequently recognized as major drivers in several types of cancers [35]. Synergy between epigenetic and genetic drivers may be particularly important in breast cancer considering the essential role of ER $\alpha$  in these malignancies and the intimate crosstalk between ER $\alpha$  and the chromatin [21, 22, 36]. For example, it is possible that PBX1 overexpression may lead to an increased number of chromatin accessible sites thereby allowing ER $\alpha$  binding near oncogenes [37]. While several other pioneer factors play a central role in modulating ER $\alpha$  binding in breast cancer cells [16, 22, 36, 38, 39], our data suggest that PBX1 has some unique properties. For example, FOXA1 [39] and GATA3 [40] play a central role in mammary and ductal morphogenesis and their interaction with ER $\alpha$  can be regarded as developmentally conserved [41, 42]. On the other hand PBX1 expression is restricted to ER $\alpha$  transformed cells and practically absent in normal ductal cells. In addition, FOXA1 and GATA3 are frequently targeted by mutations in breast cancer [43] (Fig. S11), while PBX1 is rarely affected by somatic mutations (V117M and H425N, 2 out of 892 patients). Furthermore FOXA1 and GATA3 are frequently downregulated in basal tumour but rarely over-expressed or amplified in luminal cancer (Fig. S11) [44] while PBX1 is among the top upregulated genes and is frequently amplified in luminal cancers (Figs. 4–5, S6, S8, S10 and S11).

Despite the impossibility to elucidate the temporal details of PBX1 amplification/overexpression, our data strongly suggest that PBX1 upregulation may contribute to breast cancer progression and metastatic development. Clonal evolution is thought to contribute significantly to cancer progression [45]. In this context, we can envision two possible models: in the first, cells carrying PBX1 amplification may be more adapted to escape the primary site and colonize distal tissues. On the other hand, it is possible that endocrine therapies exert positive selective pressure on cells that incur in Chr1q23.3 gains. In both scenarios, it is tempting to speculate that PBX1 would contribute to epigenetic reprogramming by interpreting or shaping the histone modifications landscape [21]. Ultimately, these changes would modulate ER $\alpha$  signaling and allow breast cancer cells to have an increased fitness in response to a diverse array of signalling pathways.

## MATERIALS AND METHODS

### Cell lines

MCF-7 and T47D cells were cultured and maintained as previously described [16]. EGF stimulation was performed as previously described [15]. Proliferation curves were performed as previously described [16]. Treatment with GSI (10 $\mu$ M of MRK003 or PF03084014) was performed as previously described [5]. Experiments were repeated independently at least five times.

### RT-qPCR and microarray

RNA samples from siControl or siPBX1 treated MCF7 in the presence or absence of EGF were extracted with RNAeasy (Quiagen) and retro-transcribed with iSCRIPT (Biorad). For microarrays, RNAs were hybridized on HT12 human beads array (Illumina Inc.). Analyses were performed using BRB- Array Tools Version 3.8.1. Raw intensity data were log<sub>2</sub> transformed, median normalized and filtered to remove non- detected spots as determined by Illumina Software. The normalization was performed by computing a gene-by-gene difference between each array and the median (e.g. reference) array, and subtracting the median difference from the log intensities on that array, so that the gene-by-gene difference between the normalized array and the reference array is zero. Two class non-paired comparison analyses were performed by computing a *t*-test for each gene using normalized log-intensities. Differentially expressed genes were determined at a significance level of *q* less than 0.01. A four class ANOVA at *p* less than 0.01 was also performed to identify genes expressed differentially across the four groups. Hierarchical clustering was employed by calculating Euclidean distance to generate heat maps for subsets of significant genes using the open source software Cluster/Treeview.

### ChIP and ChIP-seq

ChIP assay were conducted as described previously described using PBX1 (Abnova, M01 clone 4A2) or ER $\alpha$  (Santa Cruz biotechnology) [16] using 5–10\* 10<sup>6</sup> cells. Library preparation for next-generation sequencing was performed according to manufacturer's instruction starting with 5 ng of material (Illumina Inc.). Single paired libraries were sequenced using the HT2500 (Illumina Inc). Over 20 million 50 bp reads were generated through for the ER $\alpha$  ChIP and Input samples. Of those, over 92% were aligned to the human reference genome. These reads were aligned using bowtie [46] (default setting). The MACS peak-calling algorithm was used to call significantly enriched peaks using default settings (*P* < 10<sup>-5</sup>) [47].

### Ontology analysis and heatmap analysis

Ontology analyses were performed using GREAT [48]. Heatmaps were generated using CHASE [49].

### Imperial tissue microarray of paired primary and secondary breast cancers

Twenty primary breast carcinomas with a paired metastasis were acquired from the pathology archives of Charing Cross Hospital, London, UK. A tissue microarray was constructed using a manual microarrayer (Beecher) and 0.6 mm punches. The tissue microarray was immunohistochemically profiled for PBX1 and other biomarkers as previously described [16]. PBX1 monoclonal antibody (Abnova, (M01), clone 4A2) was optimized to a working concentration using tissue sections (5  $\mu$ m). Antigen retrieval was performed using 10mM Tris-HCl with 1 mM EDTA (pH 9.0), heated in a water bath at 95°C for 30 min followed by cooling at room temp for 30 min. Blocking was done using 0.3% hydrogen peroxide in PBS, followed by normal goat serum (20  $\mu$ l per ml) for 30 min. The primary antibody (1:50 in PBS) was incubated overnight at 4°C, then detected using anti-rabbit secondary antibody, Vectastain Elite peroxidase ABC kit, and ImmPACT DAB kit (all from Vector Laboratories). Subsequently, immunohistochemistry was performed on TMA sections (4–5  $\mu$ m) using the optimized staining protocol, including negative controls (omission of the primary antibody). PBX1 immunoreactivity was localized in the nucleus. Staining was scored based on the H-score and Allred Quick score (LM, DKP and PV).

### Nottingham tenovus primary breast cancer series

Primary operable breast cancer cases (*n* = 1650) from the Nottingham Tenovus Primary Breast Carcinoma Series were used and were utilized for immunohistochemistry. Clinical data were maintained on a prospective basis with a median follow-up of 126 months [23, 50]. The tissue microarrays and full-face sections from the Nottingham Tenovus Primary breast cancer series were immunohistochemically profiled for PBX1 and other biological antibodies as previously described. PBX1 mouse monoclonal antibody (Abnova, (M01), clone 4A2 [16] was optimized to a working concentration, utilizing  $\mu$ m matched full- face excisional tissue sections. Antigen retrieval was performed using Leica ER2 (pH 9.0) retrieval solution, water bath at 95C for 35 minutes followed by TBS at 50C for 10minutes. Blocking was done using Thermo Fisher UltraV block 5mins. Antibody concentration was 1:50 overnight at 4°C. Subsequently, 4  $\mu$ m TMA sections were immunostained using the optimized staining protocol. Detection

was achieved using the Novalink Polymer Detection kit (Leica Micro- systems Inc., USA). Negative controls were performed by omission of the primary antibody. IHC revealed that PBX1 had a nuclear location (4B). Nuclear staining was scored based on the H-score and Allred Quick score. Determination of the optimal cut-offs was performed using histograms and confirmed using X-tile bio-informatics software (Yale University, USA) [51] where data were split into training and validation sets. A total of 1198 tumours were suitable for analysis (460 PBX1 negative and 738 PBX1 positive). The tumour cores were evaluated by two of the coauthors (DKP and PV) blinded to the clinico-pathological characteristics of patients. There was substantial intra- and inter- observer agreement (k [0.768; Cohen's j and multi-rater j tests, respectively).

### Survival analysis using transcriptomics data

Positive PBX1 expression was dichotomised to 90% using distant metastases as an outcome and determined through X-tile software (University of Yale, Yale, USA) [51]. Prognostic analyses were undertaken using Kaplan-Meier curves using the Log-rank test. A *p*-value < 0.05 was considered significant.

### Survival analysis using transcriptomics data

Database construction and survival analysis was performed as described previously [52]. For the expression of the genes, the median expression was used as the cut-off in a Cox regression analysis. Kaplan-Meier survival plot, and hazard ratio with 95% confidence intervals and logrank *P* value were calculated and plotted in R using Bioconductor packages.

### Construction of METABRIC microarray database

The Metabric project employed Illumina microarrays to measure gene expression across all genes. The raw gene chip files were downloaded from the European Genome-phenome Archive (EGA) (<https://www.ebi.ac.uk/ega/>) [25]. The entire cohort contains 1988 samples, the average overall survival is 8.07 years, 76% of the samples are ER positive and 47.3% are node positive. Due to batch effects between the Metabric training and validation sets we have not used the pre-normalized tables but have re-run the complete pre-processing for all arrays. In this, raw data was first imported into R and summarized using the beadarray package [53]. For annotation, the Illumina Humanv3 database of Bioconductor was used (<http://www.bioconductor.org>). Finally, quantile normalization was performed using the preprocessCore package [54].

### FAIRE and DNaseI-hypersensitivity-seq

FAIRE-seq and DHS-seq analysis were performed as previously described [16, 55] with minor modifications.

### Circulating free DNA analysis

DNA was extracted from blood cell pellets and 1ml plasma as described previously [30]. To confirm amplification of PBX1, 10ng cfDNA was subjected to 5 cycles of preamplification with a mix of primers (including PBZ1 and reference genes). Each sample was then analysed in triplicate by real-time quantitative PCR (qPCR) in a 10 µl reaction volume. Reactions were run on an Applied Biosystems thermal cycler (Step One Plus) and analysed with Step One v2.1 software and Microsoft Excel.

### Primers

Primers sequences (RT-qPCR, ChIP-qPCR) are available upon requests. Primers to study PBX1 amplification were as follow: Primers PBX1-F: AGCCCACTCATCTTACGTGAC PBX1-R: ACGAAATTCCACTCCA ACTCCA Probe PBX1 GCTCAGGCCTATCTTCTGGA FAM-MGB.

### ACKNOWLEDGMENTS AND FUNDING

This work was supported by an Imperial Junior Research Fellowship and a Breast Cancer Campaign Pilot Grant (P46869) to LM, Wellcome Trust [103034/Z/13/Z] to DKP and the OTKA K108655 grant to BG. The Imperial College Healthcare NHS Trust Tissue Bank provided tissue samples. Other investigators may have received samples from these same tissues' The research [was funded / supported by] the National Institute for Health Research (NIHR) Biomedical Research Centre based and Experimental Cancer Medicine Centre, Cancer Research UK Centre at Imperial College Healthcare NHS Trust Imperial College London. M.L. holds a young investigator award from the Ontario Institute for Cancer Research (OICR) and a new investigator salary award from the Canadian Institute of Health Research (CIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

### CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

### Authors' contribution

LM designed the study. LM, EWL, BG, ML and CC performed bioinformatic analysis and interpreted the data. LM, BG, ML and CC wrote the manuscript. LM, DP, VN,

SPH, DFG, YL, MF and ARG performed lab experiments. LM, DP, MF, JHS, NP, PV, AG, CN, IOE, SS and EAR performed and interpreted the pathological analysis. CC, LW, KP, DG and LP recruited patients and collected and prepared clinical specimens. LM, DFG and JS performed and analyzed cfDNA data.

## Abbreviations

ER $\alpha$ , Estrogen Receptor  $\alpha$ ; EGF, Epidermal Growth Factor; TF, Transcription factor; cfDNA, circulating free-DNA FAIRE, Formaldehyde assisted isolation of regulatory elements; CNV, copy number variation.

## REFERENCES

1. Early Breast Cancer Trialists Collaborative Group EBCTCG Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P, Pan HC, Taylor C, Wang YC, Dowsett M, Ingle J, Peto R. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011; 378:771–84.
2. Tam WL, Weinberg RA. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med*. Nature Publishing Group 2013; 19:1438–49.
3. You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell*. 2012; 22:9–20.
4. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. *Cell*. 2011; 147:275–92.
5. Magnani L, Stoeck A, Zhang X, Lanczky A, Mirabella AC, Wang TL, Györfy B, Lupien M. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc Natl Acad Sci USA*. 2013.
6. Di Leo A, Jerusalem G, Petruzelka L, Torres R, Bondarenko IN, Khasanov R, Verhoeven D, Pedrini JL, Smirnova I, Lichinitser MR, Pendergrass K, Garnett S, Lindemann JPO, Sapunar F, Martin M. Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer. *J Clin Oncol*. 2010; 28:4594–600.
7. Pery L, Paridaens R, Hawle H, Zaman K, Nolé F, Wildiers H, Fiche M, Dietrich D, Clément P, Köberle D, Goldhirsch A, Thürlimann B. Clinical benefit of fulvestrant in postmenopausal women with advanced breast cancer and primary or acquired resistance to aromatase inhibitors: final results of phase II Swiss Group for Clinical Cancer Research Trial (SAKK 21/00). *Ann Oncol*. 2007; 18:64–9.
8. Chia S, Gradishar W, Mauriac L, Bines J, Amant F, Federico M, Fein L, Romieu G, Buzdar A, Robertson JFR, Brufsky A, Possinger K, Rennie P, Sapunar F, Lowe E, Piccart M. Double-blind, randomized placebo controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: results from EFACT. *J Clin Oncol*. 2008; 26:1664–70.
9. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, Li Z, Gala K, Fanning S, King TA, Hudis C, Chen D, Taran T, Hortobagyi G, Greene G, Berger M, Baselga J, Chandralapaty S. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet*. 2013; 45:1439–45.
10. Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L, Gursky A, Siddiqui J, Tomlins SA, Roychowdhury S, Pienta KJ, Kim SY, Roberts JS, Rae JM, Van Poznak CH, Hayes DF, Chugh R, Kunju LP, Talpaz M, Schott AF, Chinnaiyan AM. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet*. 2013; 45:1446–51.
11. Barone I, Iacopetta D, Covington KR, Cui Y, Tsimelzon A, Beyer A, Andò S, Fuqua SAW. Phosphorylation of the mutant K303R estrogen receptor alpha at serine 305 affects aromatase inhibitor sensitivity. *Oncogene*. 2010; 29:2404–14.
12. Lupien M, Jeyakumar M, Hébert E, Hilmi K, Cotnoir-White D, Loch C, Auger A, Dayan G, Pinard G-A, Wurtz JM, Moras D, Katzenellenbogen J, Mader S. Raloxifene and ICI182, 780 increase estrogen receptor-alpha association with a nuclear compartment via overlapping sets of hydrophobic amino acids in activation function 2 helix 12. *Mol Endocrinol*. 2007; 21:797–816.
13. Bunone G, Briand PA, Miksicek RJ, Picard D. Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J*. 1996; 15:2174–83.
14. Osborne CK, Schiff R. Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer. *Breast*. 2003; 12:362–7.
15. Lupien M, Meyer CA, Bailey ST, Eeckhoutte J, Cook J, Westerling T, Zhang X, Carroll JS, Rhodes DR, Liu XS, Brown M. Growth factor stimulation induces a distinct ER(alpha) cistrome underlying breast cancer endocrine resistance. *Genes Dev*. 2010; 24:2219–27.
16. Magnani L, Ballantyne EB, Zhang X, Lupien M. PBX1 genomic pioneer function drives ER $\alpha$  signaling underlying progression in breast cancer. *PLoS Genet*. 2011; 7:e1002368.
17. Park JT, Shih I-M, Wang T-L. Identification of Pbx1, a potential oncogene, as a Notch3 target gene in ovarian cancer. *Cancer Res*. 2008; 68:8852–60.
18. Martin L-A, Farmer I, Johnston SRD, Ali S, Dowsett M. Elevated ERK1/ERK2/estrogen receptor cross-talk enhances estrogen-mediated signaling during long-term estrogen deprivation. *Endocr Relat Cancer*. 2005; 12:S75–84.

19. Tyson JJ, Baumann WT, Chen C, Verdugo A, Tavassoly I, Wang Y, Weiner LM, Clarke R. Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. *Nat Rev Cancer*. 2011; 11:523–32.
20. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet*. 2010; 43:27–33.
21. Magnani L, Eeckhoutte J, Lupien M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet*. 2011; 27:465–74.
22. Jozwik KM, Carroll JS. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer*. 2012; 12:381–5.
23. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991; 19:403–10.
24. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61–70.
25. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012; 486:346–52.
26. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2:401–4.
27. Sarajlić A, Filipović A, Janjić V, Coombes RC, Pržulj N. The role of genes co-amplified with nicastrin in breast invasive carcinoma. *Breast Cancer Res Treat*. 2014; 143:393–401.
28. Page K, Hava N, Ward B, Brown J, Guttery DS, Ruangpratheep C, Blighe K, Sharma A, Walker RA, Coombes RC, Shaw JA. Detection of HER2 amplification in circulating free DNA in patients with breast cancer. *Br J Cancer*. 2011; 104:1342–8.
29. Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, Foy CA, Scott DJ. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Res*. 2012; 40:e82.
30. Shaw JA, Page K, Blighe K, Hava N, Guttery D, Ward B, Brown J, Ruangpratheep C, Stebbing J, Payne R, Palmieri C, Cleator S, Walker RA, Coombes RC. Genomic analysis of circulating cell free DNA infers breast cancer dormancy. 2011.
31. Umetani M, Domoto H, Gormley AK, Yuhanna IS, Cummins CL, Javitt NB, Korach KS, Shaul PW, Mangelsdorf DJ. 27-Hydroxycholesterol is an endogenous SERM that inhibits the cardiovascular effects of estrogen. *Nat Med*. 2007; 13:1185–92.
32. Blair RM, Fang H, Branham WS, Hass BS, Dial SL, Moland CL, Tong W, Shi L, Perkins R, Sheehan DM. The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. *Toxicol Sci*. 2000; 54:138–53.
33. Nelson ER, Wardell SE, Jasper JS, Park S, Suchindran S, Howe MK, Carver NJ, Pillai RV, Sullivan PM, Sondhi V, Umetani M, Geradts J, McDonnell DP. 27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology. *Science*. 2013; 342:1094–8.
34. Wu Q, Ishikawa T, Sirianni R, Tang H, McDonald JG, Yuhanna IS, Thompson B, Girard L, Mineo C, Brekken RA, Umetani M, Euhus DM, Xie Y, Shaul PW. 27-Hydroxycholesterol promotes cell-autonomous, ER-positive breast cancer growth. *Cell Rep*. 2013; 5:637–45.
35. Suva ML, Riggi N, Bernstein BE. Epigenetic Reprogramming in Cancer. *Science*. 2013; 339:1567–70.
36. Magnani L, Lupien M. Chromatin and Epigenetic Determinants of Estrogen Receptor Alpha (ESR1) Signaling. *Mol Cell Endocrinol*. 2013.
37. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR, Ali S, Chin S-F, Palmieri C, Caldas C, Carroll JS. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012; 481:389–93.
38. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*. 2010; 143:470–84.
39. Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008; 132:958–70.
40. Theodorou V, Stark R, Menon S, Carroll JS. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res*. 2012.
41. Chou J, Provot S, Werb Z. GATA3 in development and cancer differentiation: cells GATA have it! *J Cell Physiol*. 2010; 222:42–9.
42. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, Godwin AK, Korach KS, Visvader JE, Kaestner KH, Abdul-Karim FW, Montano MM, Keri RA. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development*. 2010; 137:2045–54.

43. Polyak K, Metzger Filho O. SnapShot: Breast Cancer. *Cancer Cell*. 2012; 22:562–562.e1.
44. Brunner AL, Li J, Guo X, Sweeney RT, Varma S, Zhu SX, Li R, Tibshirani R, West RB. A shared transcriptional program in early breast neoplasias despite genetic and clinical distinctions. *Genome Biol*. 2014; 15:R71.
45. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012; 481:306–13.
46. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25.
47. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9:R137.
48. McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010; 28:495–501.
49. Nielsen CB, Younesy H, O’Geen H, Xu X, Jackson AR, Milosavljevic A, Wang T, Costello JF, Hirst M, Farnham PJ, Jones SJM. Spark: a navigational paradigm for genomic data exploration. *Genome Res*. 2012; 22:2262–9.
50. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology*. 1992; 20:479–89.
51. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bioinformatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res*. 2004; 10:7252–9.
52. Györfy B, Lanczky A, Szallasi Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr Relat Cancer*. 2012; 19:197–208.
53. Dunning MJ, Smith ML, Ritchie ME, Tavaré S. *beadarray: R classes and methods for Illumina bead-based data*. *Bioinformatics*. (Oxford, England): 2007; 23:2183–4.
54. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. Oxford, England: 2003; 19:185–93.
55. Ling G, Sugathan A, Mazor T, Fraenkel E, Waxman DJ. Unbiased, genome-wide *in vivo* mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression. *Mol Cell Biol*. 2010; 30:5531–44.