

Neoadjuvant tamoxifen synchronizes ER α binding and gene expression profiles related to outcome and proliferation

Tesa M. Severson¹, Ekaterina Nevedomskaya^{1,2}, Justine Peeters³, Thomas Kuilman⁴, Oscar Krijgsman⁴, Annelot van Rossum¹, Marjolein Droog¹, Yongsoo Kim^{1,2}, Rutger Koornstra⁵, Inès Beumer³, Annuska M. Glas³, Daniel Peeper⁴, Jelle Wesseling¹, Iris M. Simon³, Lodewyk Wessels², Sabine C. Linn^{1,6,7} and Wilbert Zwart¹

¹ Division of Molecular Pathology, Netherlands Cancer Institute, Amsterdam, CX, The Netherlands

² Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Amsterdam, CX, The Netherlands

³ Agendia NV, Amsterdam, XH, The Netherlands

⁴ Division of Molecular Oncology, Netherlands Cancer Institute, Amsterdam, CX, The Netherlands

⁵ Department of Medical Oncology, Radboud University Medical Center, Nijmegen, GA, The Netherlands

⁶ Division of Medical Oncology, Netherlands Cancer Institute, Amsterdam, CX, The Netherlands

⁷ Department of Pathology, University Medical Center Utrecht, CX, The Netherlands

Correspondence to: Wilbert Zwart, **email:** w.zwart@nki.nl

Sabine C. Linn, **email:** s.linn@nki.nl

Keywords: ChIP-seq, estrogen receptor, endocrine therapy, neoadjuvant therapy, gene expression analysis

Received: April 07, 2016

Accepted: April 13, 2016

Published: April 25, 2016

ABSTRACT

Estrogen receptor alpha (ER α)-positive breast cancers are frequently treated with tamoxifen, but resistance is common. It remains elusive how tamoxifen resistance occurs and predictive biomarkers for treatment outcome are needed. Because most biomarker discovery studies are performed using pre-treatment surgical resections, the effects of tamoxifen therapy directly on the tumor cell *in vivo* remain unexamined. In this study, we assessed DNA copy number, gene expression profiles and ER α /chromatin binding landscapes on breast tumor specimens, both before and after neoadjuvant tamoxifen treatment. We observed neoadjuvant tamoxifen treatment synchronized ER α /chromatin interactions and downstream gene expression, indicating that hormonal therapy reduces inter-tumor molecular variability. ER α -synchronized sites are associated with dynamic FOXA1 action at these sites, which is under control of growth factor signaling. Genes associated with tamoxifen-synchronized sites are capable of differentiating patients for tamoxifen benefit. Due to the direct effects of therapeutics on ER α behavior and transcriptional output, our study highlights the added value of biomarker discovery studies after neoadjuvant drug exposure.

INTRODUCTION

Breast cancer is the most common cancer among women in the world. In 2012 over 1.6 million women were diagnosed with breast cancer worldwide. In the same year, around 520,000 women died from breast cancer [1]. The inter-patient heterogeneous nature of breast cancer is evident and clinically relevant histological and

molecular subtypes of breast cancer can be identified. [2-9]. The major subtype—estrogen receptor alpha (ER α)-positive luminal breast cancer—is clinically defined by nucleic protein levels of ER α [10]. ER α is a ligand-dependent transcription factor, activated by the natural hormone estradiol. Ligand-bound ER α can bind the DNA and recruit co-factors that form the foundations of the transcription complex, ultimately affecting target gene expression and driving tumor cell proliferation [11]. ER α -

positive breast cancer has specific clinicopathological characteristics including a favorable prognosis compared with ER α negative disease, particularly in the first few years following diagnosis [12] and more often low tumor grade in comparison with other breast cancer subtypes [4]. Patients with ER α -positive breast cancer make up over 70% of the total breast cancer population worldwide [13].

Since ER α is the major driver in luminal breast cancers, endocrine therapies have been developed to limit its transcriptional potency, including tamoxifen and aromatase inhibitors [14]. Tamoxifen is aimed to competitively bind the ER α , which prevents co-factor recruitment and disrupts the transcriptional complex formation [15]. Through this mechanism, tamoxifen blocks ER α -responsive gene expression and inhibits ER α -positive tumor cell proliferation [16]. Aromatase inhibitors inhibit the production of estrogen and consequently inhibit tumor cell proliferation [17]. Although these endocrine therapies are successful agents, resistance to treatment is common. However, after developing a relapse despite tamoxifen treatment, 50% of patients still do respond to aromatase inhibitors [18]. Analogous to this, 50% of patients with acquired resistance to aromatase inhibitors still respond to tamoxifen [18]. This study and other reports [13] illustrate the heterogeneous response to first and second line endocrine therapy in ER α -positive breast cancer. Critically, these studies also exemplify that development of reliable predictive biomarkers for selective treatment efficacy is an urgent medical need.

From a clinical perspective, identification of patients most likely to benefit from specific endocrine treatments in the early ER α -positive breast cancer setting is paramount. Currently, biomarker discovery is guided by data (*e.g.* gene expression signatures) collected from the untreated primary tumor samples obtained during surgery. However, the prediction of longitudinal therapy response may be more reliable when data are available from samples after drug exposure. In the endocrine therapy setting, neoadjuvant studies have been used successfully to determine response [19, 20], indicating the potential of molecular data taken from post-treatment samples to gain clinical knowledge. Here, we present the first combined characterization of transcriptomic, epigenomic, genomic and clinical data analyzed from ER α -positive breast cancers both before and after 2-6 weeks of neoadjuvant tamoxifen treatment. By integrating these genomic data-streams from different time points, we aimed to examine drug-induced differences at the genetic, epigenetic and transcriptional level with the goal to identify novel predictive biomarkers for tamoxifen-response.

RESULTS

AFTER study and patient characteristics

Forty-eight ER α -positive patients were enrolled at two hospitals in the Netherlands and sample material was taken from 28 tamoxifen-treated patients for further analysis (Figure 1A, Supplemental Figure 1A, 1B). Patient characteristics and collection data can be found in Table I. Patient treatment details can be found in Supplemental Table I. We collected immunohistochemistry (IHC) data for ER α and PR for both pre- and post-treatment samples of 26 of 28 patients (Supplemental Figure 1A). No tumors were HER2-positive. Clinical marker levels (ER α and PR) did not differ between sample origin hospitals. We found no significant differences between pre- and post-tamoxifen treatment ER α levels (Wilcoxon rank-sum, $P = 0.26$) (Figure 1B, 1C). A trend was identified for higher PR levels after tamoxifen treatment, however it was not significant (Wilcoxon rank-sum, $P = 0.08$) (Figure 1B, 1C). Gene expression levels for the corresponding coding genes, Estrogen receptor 1 (ESR1) and progesterone receptor (PGR) were also not significantly different between pre- and post-treatment samples (data not shown). When investigating the cell proliferation marker Ki67 (MIB1), post-treatment samples had significantly lower gene expression (MKI67) levels (Wilcoxon rank-sum, $P < 0.01$) (Figure 1D). A similar significant trend was found when measuring Ki67 at the protein level with IHC (data not shown). Because IHC Ki67 score is difficult to interpret with substantial inter-observer variation [43], we chose to examine microarray gene expression levels of MKI67 as they are less subjective. In addition, we classified all our samples for two known molecular classifiers with links to outcome, MammaPrint and IntClust (Supplemental File 1).

Tamoxifen reprograms ER α binding in tumors

Since tamoxifen is aimed to directly target ER α action, ER α functioning may also be affected on the genomic level by drug treatment. Therefore, we analyzed the chromatin binding profiles of ER α in 6 pre- and 8 post-treatment samples (4 treatment pairs) using ER α ChIP-seq (Figure 2A; Supplemental Figure 1A). We identified ER α bound genomic regions, as exemplified for two typical ER α regions in the human genome found at the enhancer regions proximal to the RARA and IGFBP4 loci (Supplemental Figure 2A). As expected, the top DNA binding motifs identified in both pre- and post-treatment were found to be the hormone nuclear receptor family with ESR1 as the top factor (Supplemental Figure 2B).

A relocation of ER α chromatin interactions to other

sites was observed after tamoxifen treatment (Figure 2B, Supplemental File 2). Our laboratory [44] and others [26] previously reported ER α chromatin interaction profiles as highly heterogeneous between tumors. Consistent with previous findings, we found limited overlap of ER α binding patterns between two tumors prior to therapy (Figure 2C). Importantly, the pattern overlap was substantially increased after neoadjuvant tamoxifen therapy in treatment pairs (Figure 2C). To investigate this further in a quantitative fashion, the percent of overlapping ER α bound genomic regions between all samples of the total ER α bound genomic regions was calculated in each possible sample combination (within pre- or post-treatment condition) (Figure 2D). The average percent overlap in pre- and post-treatment samples increased from 3.0 to 7.9 with data ranging from 0.01% to 13% overlap among pre-treatment samples and 0.8% to 21% overlap

among post-treatment samples. We identified a significant (Wilcoxon signed rank, $P < 0.001$) increase in overlap in genomic regions in the post-treatment group compared with the pre-treatment group.

Next, we characterized the genomic regions of the tamoxifen-induced synchronized sites (I) and pre-treatment unique sites (II) (Supplemental Figure 2C). In genomic location for both, the sites are typical of what is known of ER α binding in that they are found mostly in distal intergenic and intronic regions (Supplemental Figure 2D) [45]. In addition, the most common sequence motifs associated with both groups of sites were ESR1 and ESR2 (Supplemental Figure 2E). Similar findings are observed for pre-treatment unique sites. Next, we determined the nearest gene for each of the 126 ER α bound sites (I) within 20kb [46]. Using this definition, 96 genes are associated with these tamoxifen-synchronized sites. When examining

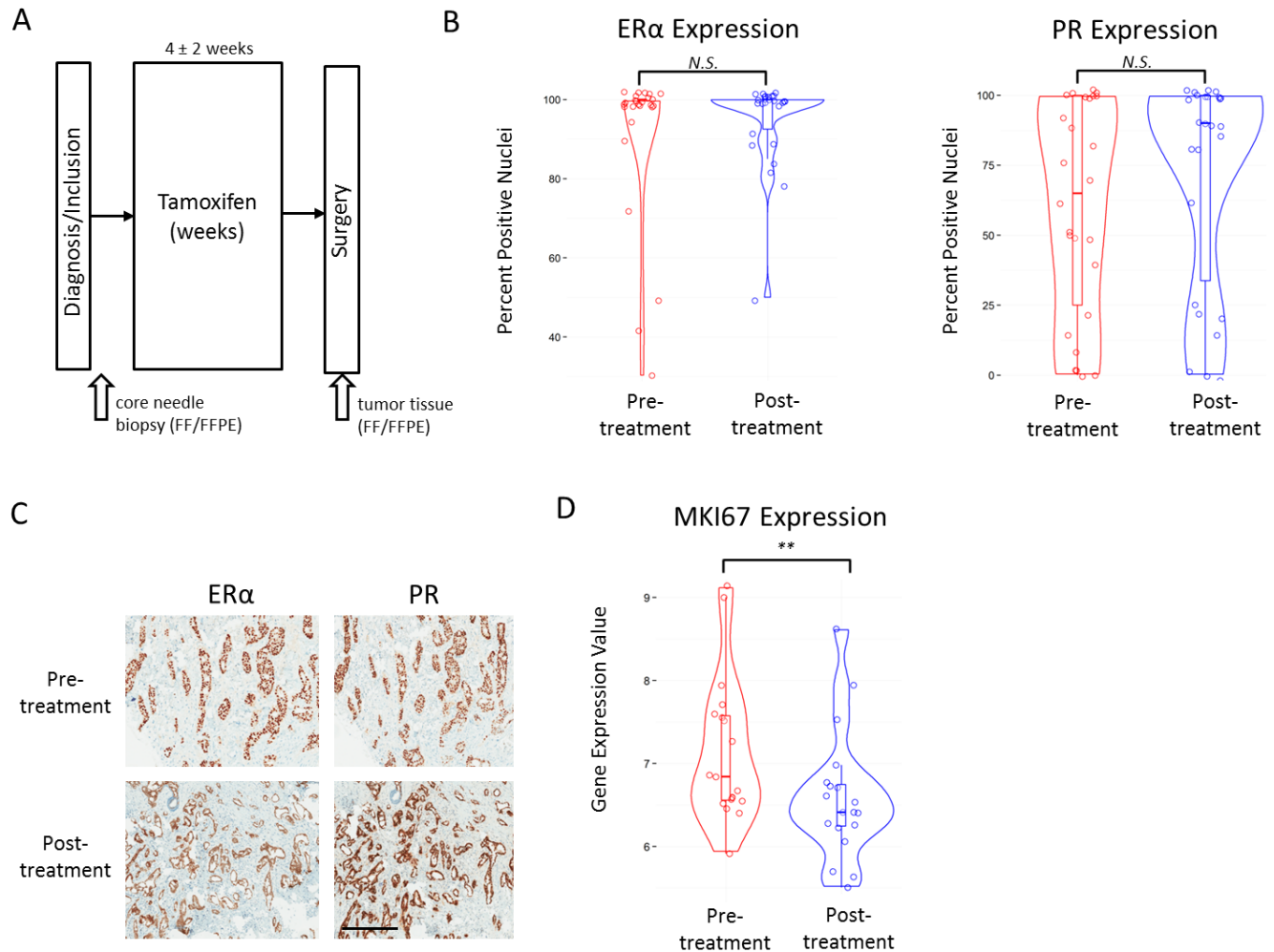


Figure 1: Study collection and clinical marker values. **A.**, Schematic representation of the AFTER study patient and material collection. **B.**, Violin plots with boxplot overlay of IHC scores (Y-axis) for ER α (left panel) and PR (right panel) in pre- and post-treatment samples. N.S. indicates a p -value that is not significant at 0.05 level. **C.**, Example of IHC staining of ER α and PR in a pre- and post-treatment sample. Black bar indicates 400 μ m. **D.**, Violin plots with boxplot overlay of MKI67 gene expression values in pre- and post-treatment samples. ** indicates $P < 0.01$.

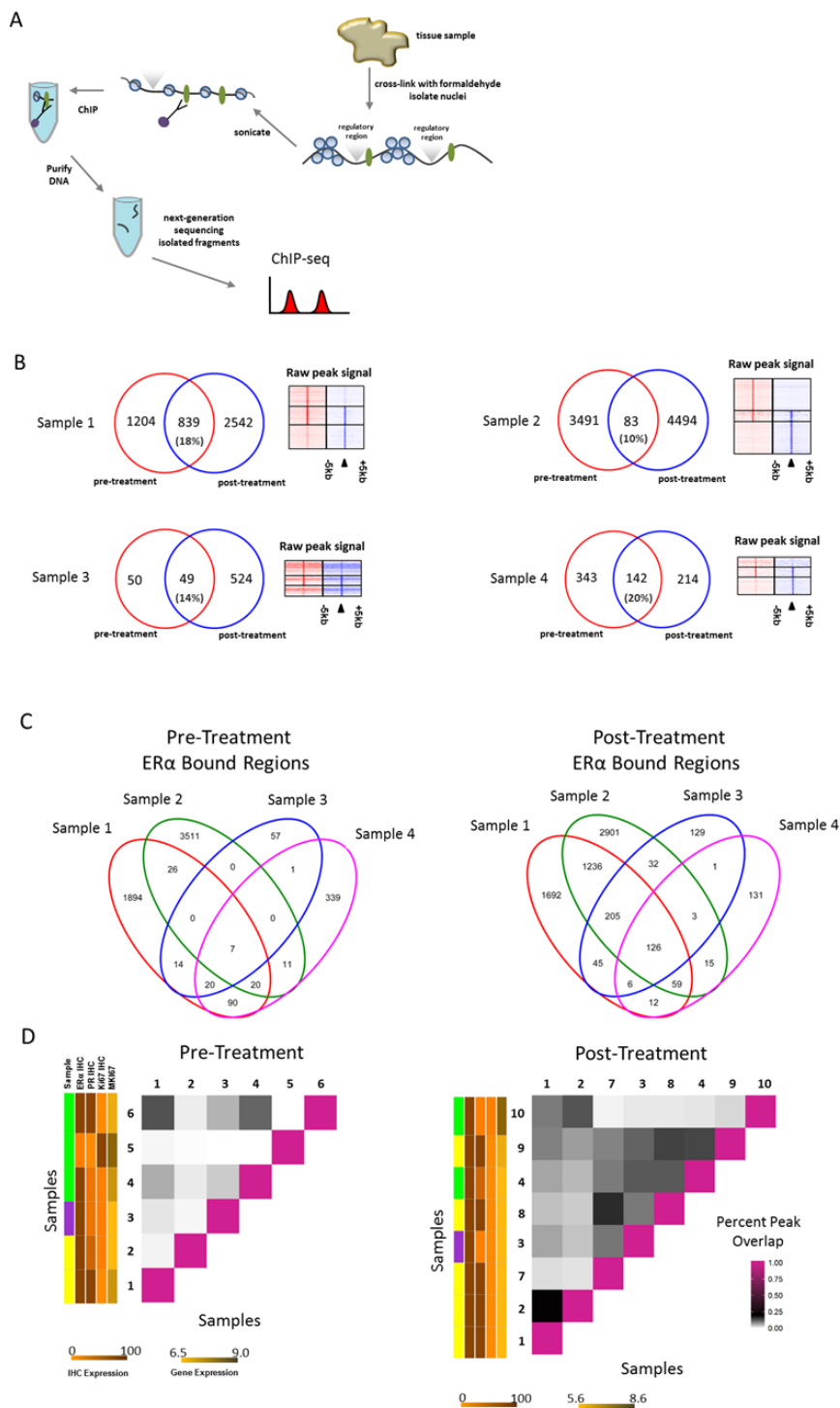


Figure 2: ER α binding events before and after neoadjuvant tamoxifen treatment. **A.**, Schematic representation of ChIP-seq process. ER α bound to the DNA is depicted as green ovals; regulatory regions are depicted above the DNA strand as grey triangles. **B.**, Venn diagrams depicting the overlap of ER α bound regions (peaks) in both pre-(red) and post-(blue) treatment pairs. Heatmaps on the right show ER α binding events at the center of the peak and the surrounding ± 5 kb for the pre-treatment sample only peaks (top), overlapping peaks (middle) and post-treatment sample only peaks (bottom). **C.**, Venn diagrams depicting all overlapping peaks for pre- and post-treatment pair samples separately. **D.**, Matrix to visualize percent of overlapping peaks (of total peaks in each pair combination) for both pre- and post-treatment samples separately. Vertical colored side bars indicate patient menopausal-status (sample) and ER α IHC, PR IHC, Ki67 IHC and MKI67 values respectively. Patient menopausal-status is shown in yellow (post-menopausal), green (pre-menopausal) and purple (male). Separate scale bars are shown below the plot for expression levels corresponding to sidebars. IHC expression scale bar for both panels indicates percentage staining nuclei. Gene expression scale bar for both panels indicates normalized gene expression values.

these genes using Ingenuity Pathway Analysis (Qiagen), we find these genes involved in drug and lipid metabolism, estrogen receptor signaling and notably, EGF was identified as a potential upstream regulator. (Supplemental Figure 2F).

ER α , PR and FOXA1 binding profiles in tumors and MCF7 cells reveal dynamic FOXA1 binding

ER α binding and activity in breast cancer is regulated and controlled by various other factors such as FOXA1, a pioneering factor required for binding of ER α [47–52] and PR, a modulator of ER α action in breast cancer [53]. When examining the 126 ER α bound sites synchronized by tamoxifen (Figure 3A (I)) in publicly available ChIP-seq data for tumors, we found the average relative signal intensity at these sites is higher in metastatic samples compared with primary tumors samples (ER α /PR positive) (Figure 3B, 3C). A similar pattern is seen in the unique pre-treatment sites (Figure 3A (II)) but with much less overall signal intensity (Figure 3B, 3C). In addition, in MCF7 cells we observed a similar pattern in response to tamoxifen or estradiol in ER α binding at these sites (Figure 4D, 4E). No PR binding was observed (Figure 3D, 3E) under full medium conditions for either the 126 synchronized sites or the unique pre-treatment sites. FOXA1 chromatin binding is described as independent of ER α action and irresponsive to hormonal stimuli [50, 51]. However, FOXA1 chromatin binding at these 126 sites was clearly induced by both tamoxifen and estradiol, accompanied by ER α binding. These data imply a unique feature of the 126 tamoxifen-synchronized ER α sites, hallmarked by dynamic FOXA1 action at these sites.

ER α , FOXA1 binding profiles in MCF7 cells reveal potential mechanistic players

To better understand a possible mechanism of ER α and FOXA1 reprogramming to these sites after addition of tamoxifen, we investigated publically available ER α and FOXA1 ChIP-seq data from MCF7 cells treated with specific mitogens and/or growth factors [50]. Data from Hurtado and colleagues indicate the overlap of ER α binding sites for both estradiol and tamoxifen treated MCF7 cells is substantial. Furthermore, the proportion of tamoxifen only binding sites is very small (7%) [50]. Within our paired tumor data, we show that on average 56.3% of the total sites are post-tamoxifen treatment only binding sites (Figure 2B).

To examine a potential mechanism involved in synchronizing ER α to these sites based on known regulators of ER α binding, we investigated the Ross-Innes mitogen cocktail data, which contains EGF, IGF-1, TNF α and IL-6 [26], the Franco TNF α data [54] and

the Lupien EGF data [52] at the tamoxifen synchronized and pre-treatment unique sites. The mitogen cocktail significantly induced FOXA1 binding at these sites (Wilcoxon signed rank, $P < 0.0001$). (Figure 4A, 4B). In ER α binding data where estradiol, TNF α or a combination of the two were added, we observed no increased FOXA1 binding by TNF α (Figure 4C, 4D). Combined, these data rule out TNF α alone as a key regulator of FOXA1 dynamics at these sites, leading us to investigate EGF as it was implicated in our own Ingenuity Pathway Analysis results. ChIP-chip data of EGF-stimulated ER α binding sites show high overlap (60%) with our tamoxifen induced sites (Figure 4E, Supplemental File 3). We found the enrichment for these sites within the pre-defined universe of known ER α binding sites to be significant (Fisher's hypergeometric test, $P < 1.10E-15$) (Supplemental File 3) indicating the high proportion of overlap of tamoxifen induced sites are biologically significant. In addition, we observed comparable results when looking at the overlap of the tamoxifen induced sites with serine 118 ER α DNA binding sites induced by estradiol (Figure 4F) [55]. Cumulatively, these data implicate EGF as a potential upstream regulator of FOXA1-induced chromatin binding at the 126 sites, facilitating ER α action at these sites in tumors post tamoxifen exposure.

DNA copy number profiling and alternate allele frequency analysis reveal little variation between replicates and/or treatment condition

We observed ER α binds to different sites after neoadjuvant tamoxifen treatment (Figure 2D). We questioned if the change in ER α binding is due to genomic changes within the tumor population or re-targeting of ER α . Notably, interaction profiles [56] and overall library read numbers [23] have been shown to be reproducible within replicate tumor samples. To determine the potential impact of intra-tumor genomic variability, we examined DNA copy number profiles (example in Supplemental Figure 3A) of 3 pairs of replicate tumor samples that were ChIP-sequenced twice. Replicate samples were sectioned from the same tumor at a different time with at least 5 cell layers (25 to 50 μ m) between replicates. DNA copy number data were obtained from the off-target read in ChIP-seq experiments using the CopywriteR method [30] (Supplemental Figure 1A) and data were subsequently segmented with circular binary segmentation. We did not identify any substantial changes in the DNA copy number profiles between sample replicates. (Supplemental Figure 3B). Unsupervised hierarchical cluster analysis on all the copy number data (20kb region bins, non-segmented) found most replicates cluster together with the exception of sample pair D ($n = 12$) (Supplemental Figure 3C). To quantify this further we calculated p -values (100 indicates

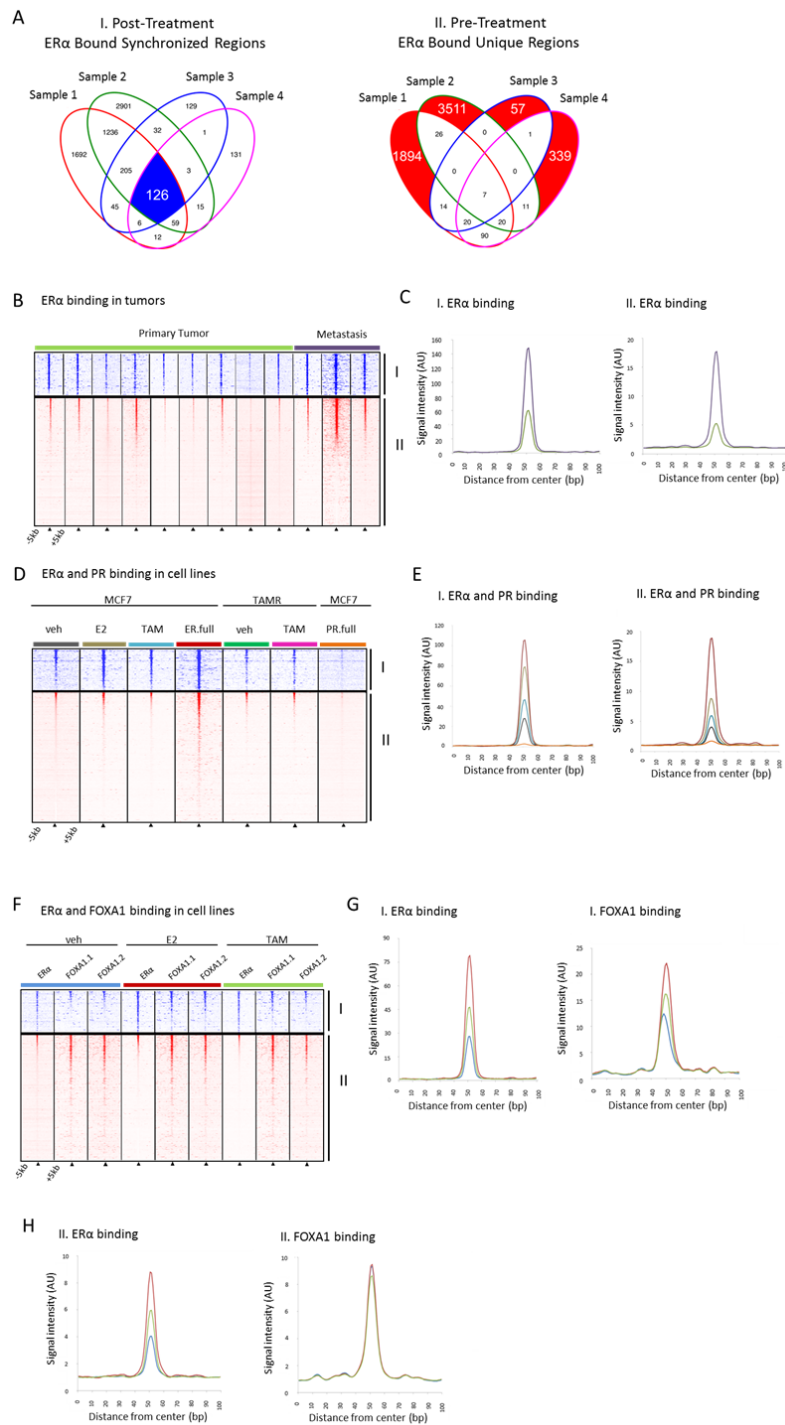


Figure 3: Binding profiles of 126 tamoxifen-synchronized regions (I) and unique pre-treatment regions (II). A., Venn diagrams depicting overlapping peaks (paired samples) for both post-treatment tamoxifen synchronized sites (I) and unique pre-treatment sites (II). B., Heatmap showing binding peak intensity for ER α binding events in I and II sites (± 5 kb) in primary tumors (green) and metastases (purple). C., Normalized average signal intensity of ER α binding events from panel B. Line colors match B. D., Heatmap showing binding peak intensity for ER α binding events in I and II sites (± 5 kb) in MCF7 cell lines deprived of hormones for three days and then given vehicle (grey), estradiol (E2), brown, tamoxifen (TAM), blue and full medium (ER.full), red. PR binding in MCF7 cell lines deprived of hormones for three days and then given full medium is shown in orange (PR.full). TAMR cell lines deprived of hormones for three days and then given vehicle (green) and TAM (purple) are also depicted. E., Normalized average signal intensity of ER α binding events from D. Line colors match D. F., Heatmap showing binding peak intensity for ER α and FOXA1 binding events in I and II sites (± 5 kb) in MCF7 cell lines deprived of hormones for three days and then given vehicle (blue), E2 (red) and TAM (green). G., H., Normalized average signal intensity of ER α (left) and FOXA1 (right) binding events from panel F. Line colors match F.

a p -value < 0.00001) for the clustering based on multiscale bootstrap resampling and bootstrap resampling using the R package 'pvclust'. (Supplemental Figure 3D) The p -values indicate the cluster analysis is highly supported by the data illustrating the low intra-tumor genomic variability of the samples as measured by DNA copy number. In addition, in our tumor ChIP-seq replicate data, we observed high correlation of read counts in known ER α binding regions [26] (Spearman's ρ 0.40 and 0.41) within replicate sample pairs and low correlation in 3 un-related samples to one of the replicate samples (Spearman's ρ , < 0.10) (Supplemental Figure 3E).

To examine whether there is evidence that neoadjuvant tamoxifen treatment selects for a specific tumor sub-clone we also analyzed the copy number profiles of 4 treatment pairs, both pre- and post-treatment (Supplemental Figure 1A). Interestingly, we found no major differences in the copy number profiles between sample timing (Supplemental Figure 3F) suggesting treatment does not affect the overall DNA makeup in our samples. Unsupervised hierarchical cluster analysis of all the copy number data (non-segmented 20kb region bins) showed pre- and post-treatment samples cluster together, supporting the conclusion that neoadjuvant tamoxifen treatment does not confer DNA copy number changes in the samples ($n = 8$) (Supplemental Figure 3F, 3G) or select for a small sub-clone with a different DNA copy number pattern. The resultant p -values indicate our cluster analysis is strongly supported by the data (Supplemental Figure 3H).

To quantitatively investigate the relative contribution of tumor DNA between treatment groups we called heterozygous single-nucleotide variants (SNVs) in both pre- and post-treatment ChIP-seq pairs. Assuming 100% tumor DNA in our samples, each heterozygous SNV should have a ratio of 50/50 for alternate/reference reads (alternate allele frequency). Changes in the tumor content between treatment conditions should be evident in alternate allele frequency. We required each SNV to have at least 10 reads covering the alternate and reference allele for analysis. A scatterplot shows the alternate and reference allele counts in each SNV called (Supplemental Figure 3I). The distribution of alternate allele frequency between treatment groups was not significantly different (Wilcoxon signed rank, $P = 0.76$). Furthermore, we observed no significant difference (Wilcoxon rank sum, $P = 0.37$) between pre- and post-treatment alternate allele frequency in paired analysis taking into account only variants identified in both cases of the treatment pair (Supplemental Figure 3J). These findings support our conclusions from unsupervised hierarchical clustering of DNA copy number that pre- and post-treatment changes are not detectable at the level of DNA.

Pre- and post-treatment differential gene expression

Since ER α is a transcriptional regulator, we next examined differences between pre- and post-treatment samples at the gene expression level for 20 pre-treatment and 26 post-treatment samples (19 treatment pairs) using microarray technology on FFPE samples (Supplemental Figure 1A). We used ANOVA analysis to determine the most differentially expressed genes between the two classes, pre- and post-treatment, and visualized the results with unsupervised hierarchical clustering (Supplemental Figure 4A). Using the top variable genes from the dataset (variance > 1 across samples) we found 189 genes to be differentially expressed (FDR < 0.001 and fold-change > 2 , Supplemental Figure 4A). To identify key biological processes regulated by the differentially expressed genes, we used Ingenuity Pathway Analysis. The most upregulated genes in post-treatment samples are found in the adipogenesis pathway and are involved in LXR/RXR activation and include FBJ murine osteosarcoma viral oncogene homolog (FOS), nuclear receptor subfamily 4, group A, member 1 (NR4A1; NUR77) and dual specificity phosphatase 1 (DUSP1) (Supplemental Figure 4B, Supplemental File 4). Diseases such as cancer and neurological/cardiovascular disease along with biological functions such as cellular movement and apoptosis were associated closely with the genes upregulated after treatment (Supplemental Figure 4B). Interestingly, the most downregulated genes in post-treatment samples compared with pre-treatment samples were associated with immune response, such as REX1 RNA exonuclease 1 homolog (REXO1L1P) (Supplemental Figure 4B, Supplemental File 4). Genes that were downregulated after treatment were also associated with diseases such as cancer, endocrine system disorders and breast carcinoma.

In addition, we found MKI67 gene expression was significantly reduced in the post-treatment samples (Wilcoxon rank-sum, $P = 0.003$) (Figure 1D). To examine the relationship between MKI67 levels and proliferation we determined the gene expression proliferation module scores based on published modules AURKA and CIN70 [34] and calculated the percent change as described in the module score between pre- and post-treatment. The percent change in known proliferation modules is highly correlated with percent change in MKI67 in our cohort (Supplemental Figure 4C, 4D). Notably, we found the pre-treatment gene expression of proliferation module genes was significantly more variable than in the post-treatment condition (Wilcoxon rank-sum, $P < 0.001$), suggesting expression changes in proliferation module genes are driven by the tamoxifen treatment (Supplemental Figure 4E).

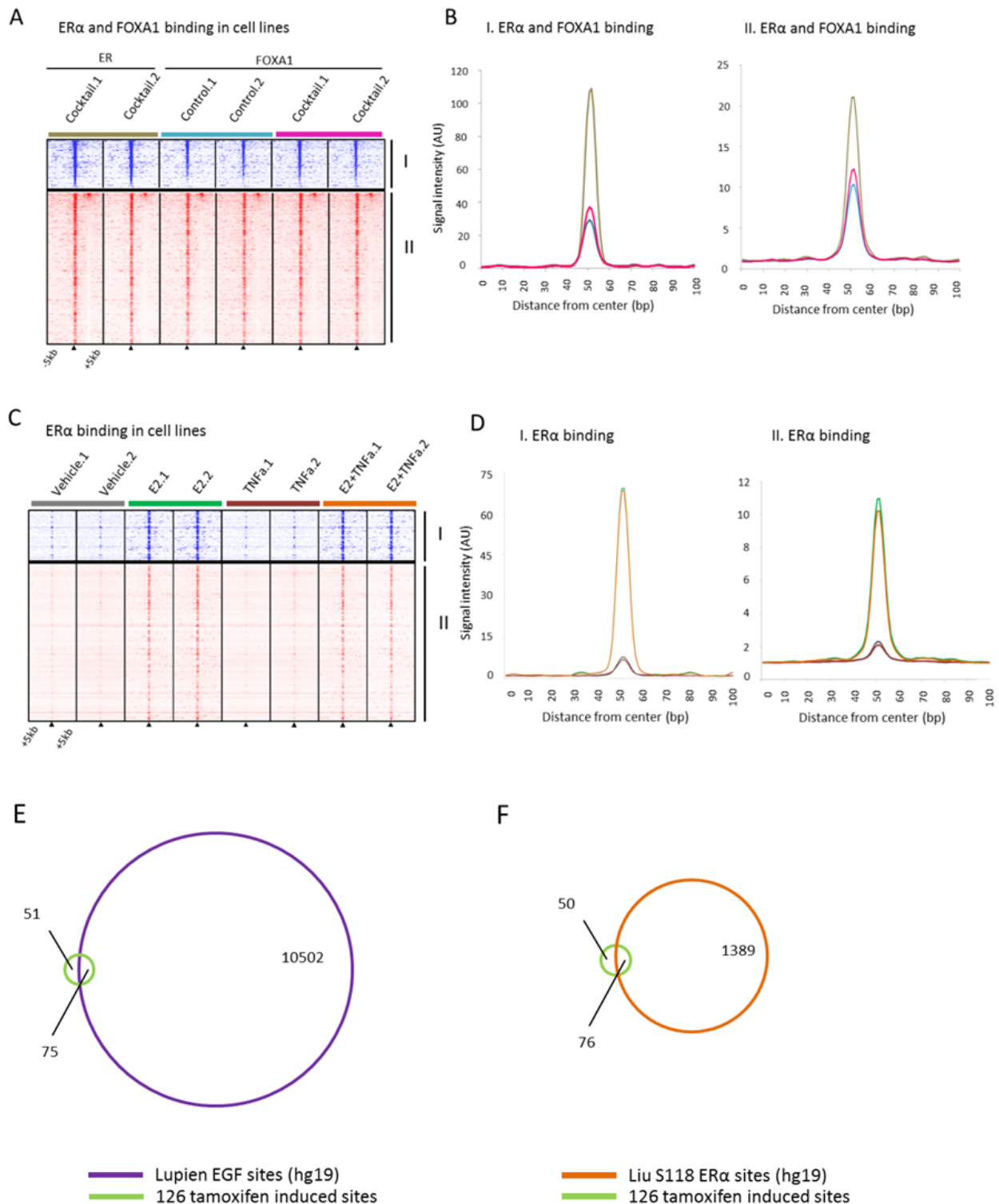


Figure 4: Binding profiles of 126 tamoxifen synchronized regions (I) and unique pre-treatment regions (II) under mitogen conditions. A., Heatmap showing binding peak intensity for ER α /FOXA1 binding events in I and II sites (\pm 5kb) in MCF7 cells, cultured under control conditions (blue) or in presence of a mitogen cocktail containing EGF, IGF-1, TNF α and IL-6 (brown for ER α and pink for FOXA1). B., Normalized average signal intensity of ER α /FOXA1 binding events from panel B. Line colors match B. C., Heatmap showing binding peak intensity for ER α binding events in I and II sites (\pm 5kb) in MCF7 cell lines deprived of hormones for three days and then given vehicle (brown), E2 (green), TNF α (burgundy), or the combination of both (orange). D., Normalized average signal intensity of ER α binding events from panel C. Line colors match D. E., Venn diagram depicting overlap of the 126 tamoxifen-induced bindings sites with EGF-induced ER α sites. F., Venn diagram depicting the overlap of tamoxifen-induced binding sites with estradiol induced S118-ER sites.

Genes associated with 126 tamoxifen induced synchronized sites differentiate patient outcome

After identifying the 96 genes associated with the 126 tamoxifen-induced sites (± 20 kb from the nearest transcription start site) (Figure 5A), we first determined these genes strongly separate pre- and post-treatment samples (Figure 5B). Next, we wished to examine their capacity to predict patient outcome using distant metastasis free survival (DMFS). For this, we used publically available gene expression and clinical data from two data sources, (i) tamoxifen treated [38, 39], $n = 250$, $n = 134$, respectively (ii) and non-endocrine therapy treated [40, 41], $n = 209$, $n = 158$, respectively (Figure 5A, 5B). For tamoxifen-treated patients ($n = 250$, [38]), the 96 genes are capable of differentiating good *versus* poor outcome for the unsupervised hierarchically defined groups using DMFS (hazard ratio (HR) = 0.49, $P = 0.006$). This result was validated in a second cohort (HR = 0.44, $P = 0.003$; $n = 134$, [39]). Based on these 96 genes, no significant differential outcome was found in the two ER α positive cohorts of non-endocrine therapy treated patients [40, 41] (HR = 1.01, $P = 0.961$; HR = 1.69, $P = 0.385$, respectively), indicating the genes are not prognostic in nature (Figure 5B).

DISCUSSION

In breast cancer, biomarker discovery is classically performed on treatment-naïve tissue samples from surgery. While this is a logical choice from a clinical perspective, it may not be the most ideal setting for predicting response to therapy. This notion was stressed by the recent discovery of specific ESR1 (the gene that encodes for ER α) mutations in relapsed tumor tissue, after having received adjuvant tamoxifen or aromatase inhibitor therapy [57, 58]. These findings illustrate that ESR1 mutations do occur in breast cancer and that, more importantly, clinically relevant data may only be discovered when the clinicopathological or molecular assessment is performed after treatment pressure has occurred.

Inter-tumor heterogeneity of breast cancer, both histopathological and molecular, is well documented [29] and recent findings indicate the tumor itself is frequently genomically heterogeneous [59, 60]. Using DNA copy number profiling we did not observe substantial differences in biological replicates from the same tumor as Wang *et al.* also previously reported using single cell whole genome sequencing [61]. When analyzing biological replicate ChIP-seq samples we found a high correlation within pairs with respect to read count in known ER α binding regions compared with un-related samples indicating the changes observed in ER α binding are induced by neoadjuvant tamoxifen treatment and not inherent to intra-tumor variability in ER α binding. In addition, we

found no evidence that treatment induces differences at the copy number or single nucleotide level suggesting that tamoxifen treatment did not select for certain sub-clones, defined by copy number aberrations and alternate allele frequency. Moreover, when examining the gene expression data alone (Mammaprint, Supplemental File 1) [7] and the integrated DNA copy number and gene expression data using IntClust classifications (Supplemental Data File 1) [36, 37] we found few changes between classes and an enrichment for the Luminal/ER α -positive subtypes. We are unable to rule out the possibility that there are differences in the DNA at the global level of single nucleotide variations with these data.

When examining the 126 tamoxifen-induced ER α bound sites in additional datasets we found a higher level of binding in metastatic samples compared with primary tumor samples from breast cancer patients. The metastatic patients from this series (Ross-Innes *et al.* 2012), failed to respond to endocrine therapy so we can categorize these samples as comparable to the post-treatment situation in our series. In this respect, our findings recapitulate the original observation of higher ER α binding site signal intensity in metastases relative to primary tumors. In addition, we found in MCF7 cells that PR binding is not present in the 126 tamoxifen synchronized sites indicating the binding at these regions is not modulated by PR. As expected, tamoxifen, estradiol and full medium conditions resulted in more binding at these sites than vehicle conditions. Interestingly, FOXA1 binding at these sites had a similar pattern in response to tamoxifen and estradiol as ER α , showing FOXA1 is dynamic at these sites while the vast majority of FOXA1 sites are not [50]. Dynamic FOXA1 sites are indicative of outcome [26] and affected by growth factors [26, 52, 54].

It has been previously reported that tamoxifen treatment induces ER α chromatin binding in MCF7 luminal breast cancer models at the same genomic sites as estradiol treatment [50]. Importantly, these data are not in line with our observation *in vivo* in 4 independent tumors treatment pairs, where tamoxifen treatment resulted in a relocation of ER α chromatin interactions to other sites. The overlap of ER α binding sites of estradiol and tamoxifen conditions for MCF7 cells is substantial [50]. We observed in our paired tumor data a far greater proportion of post-tamoxifen treatment only sites (56.3%) in contrast to what is known from cell-line data (7%). Ingenuity Pathway Analysis indicated EGF to be an upstream regulator of relocation of ER α to these sites. We found a substantial proportion of overlap of these synchronized sites with EGF-induced sites in MCF7 cells further implicating EGF as a regulator by which ER α is relocated to these sites after tamoxifen. Importantly, we also observed at these sites that FOXA1 is responsive to tamoxifen and estradiol in tumors. With this, we suggest a possible mechanism by which ER α and FOXA1 are reprogrammed to these sites after tamoxifen mediated by

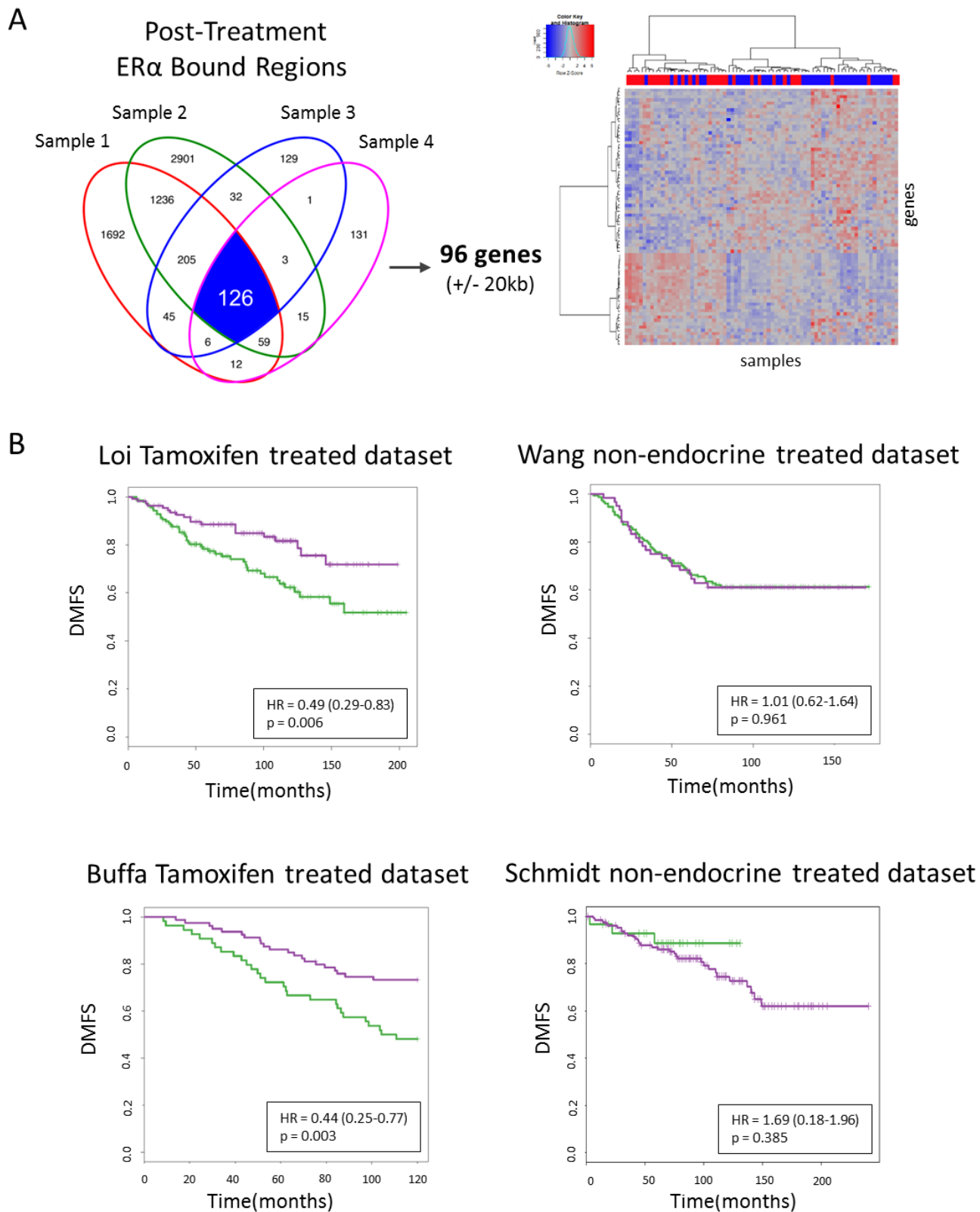


Figure 5: 96 genes survival analyses. **A.**, Venn diagram depicting overlapping peaks (paired samples) for post-treatment tamoxifen synchronized sites (I) and an unsupervised hierarchical clustering heatmap depicting gene expression in our series of those genes. Top row indicates pre-treatment samples (red) and post-treatment samples (blue). **B.**, Kaplan Meier survival curves of distant metastasis-free survival for tamoxifen treated datasets (left panels) and non-endocrine therapy treated datasets (right panels).

Table I: Patient characteristics

	Male	Premenopausal	Postmenopausal	
	<i>N</i> = 2	<i>N</i> = 14	<i>N</i> = 12	
Variable	No. (%)	No. (%)	No. (%)	<i>P</i> *
Year of diagnosis				0.004 ^a
Mean	2011	2012	2010	
Range	2011 - 2011	2011 - 2013	2008 - 2012	
Age at diagnosis				<0.001 ^a
Mean	63	47	62	
Range	52 - 73	41 - 54	52 - 79	
Treatment duration				0.189 ^a
(days)				
Mean	26	22	18	
Range	25 - 26	9 - 45	8 - 35	
Tumor size by ultrasound (mm)				0.954 ^a
Mean	20	13.6	13.8	
Range	18 - 22	0 - 27	0 - 30	
Lymph node status				0.108 ^b
micrometastasis	0 (0.0)	3 (10.7)	1 (3.6)	
positive	1 (3.6)	0 (0.0)	3 (10.7)	
negative	1 (3.6)	11 (39.3)	7 (25.0)	
other	0 (0.0)	0 (0.0)	1 (3.6)	
Tumor histological grade				0.125 ^b
1	1 (3.6)	5 (17.9)	4 (14.3)	
2	0 (0.0)	5 (17.9)	8 (28.6)	
3	1 (3.6)	4 (14.3)	0 (0.0)	
Tumor histology				0.209 ^b
IDC	2 (7.1)	11 (39.3)	11 (39.3)	
ILC	0 (0.0)	3 (10.7)	0 (0.0)	
IDC + invasive carcinoma	0 (0.0)	0 (0.0)	1 (3.6)	

*P**Tests are performed only on premenopausal and postmenopausal data; ^aWilcoxon-rank-sum-test; ^bPearson's chi-squared test

EGF. Also, our findings that FOXA1 is dynamic in these regions suggest that this specific subset of FOXA1 binding sites may be dependent on ER α .

Interestingly, our findings in both tamoxifen treated cohorts indicate genes associated with tamoxifen-induced ER α -synchronization are capable of specifically identifying breast cancer patients, who may not benefit from tamoxifen treatment. Furthermore, the observations from non-endocrine therapy treated cohorts suggest these genes are not associated with prognosis. These data are highly clinically relevant as around 50% of patients that receive tamoxifen experience a recurrence. With the 96-gene classifier reported here, 57% and 63% of patients are predicted to be tamoxifen resistant in the two tamoxifen-treated cohorts examined. Although these findings were reproducible between multiple publically available cohorts, they must be validated in the context of

a prospective randomized clinical trial in order to examine the robustness of the gene set's capacity to predict tamoxifen benefit. In addition, one would like to adjust the hazard ratios reported in this study for standard prognostic factors, such as age, histological grade *etc.* Unfortunately, for all datasets these additional clinico-pathological data are not publicly available.

A portion of ER α -positive breast cancers also express EGFR (see review by Osborne and Schiff and references therein) [62, 63]. We have observed in our own data an indication that EGF is an upstream regulator at the 126 tamoxifen-synchronized sites. From our exploration of genes associated with these sites, we determined a subset of patients do not respond well to tamoxifen treatment. These patients may benefit from a combined therapy of tamoxifen and cetuximab, an anti-EGFR antibody, or an EGFR tyrosine kinase inhibitor. Two studies in metastatic

breast cancer patients explored either tamoxifen [64] or anastrozole [65] with or without gefitinib and found a numerical advantage for the addition of gefitinib to endocrine therapy regarding clinical benefit rate and progression-free survival, but only in endocrine-therapy naïve patients, or patients who had come off adjuvant tamoxifen at least 12 months before recurrence [64, 65]. In the neoadjuvant setting postmenopausal patients with EGFR-positive, and ER α -positive breast cancer were randomized for gefitinib (EGFR tyrosine kinase inhibitor) alone or gefitinib plus anastrozole, an aromatase inhibitor. Both the single and combination agents reduced tumor size [63], and induced Ki67 downregulation as well as phospho-EGFR. Taken together, these data all support the existence of a subgroup of ER α -positive breast cancers that become endocrine therapy resistant through activation of the EGFR signaling pathway, which may be mediated by a dynamic FOXA1 DNA binding landscape. The current data will help to further define those patients that will derive benefit from the addition of an EGFR inhibitor to endocrine therapy.

Although neoadjuvant tamoxifen treatment times in this study were relatively short and varied (Table I), we did observe similar trends in transcriptomic changes in the 7 patients treated less than 2 weeks (Supplemental Figure 5A, 5B). We did not examine the DNA copy number or ChIP-seq data for these samples as only one patient in this short-treatment subgroup had paired data available. Overall, we observed a significant synchronization in the ER α /chromatin binding regions in the post-treatment samples compared with pre-treatment samples.

ER α is a key transcription factor in cellular proliferation and importantly we found gene expression proliferation module signatures are significantly less variable in the post-treatment samples indicating a synchronization of the gene expression after treatment. In addition, changes in gene expression modules are significantly correlated with changes in MKI67 levels indicating the power of MKI67 levels to detect changes in proliferation. The gene expression changes coupled with the overall stability of the DNA copy number profiles between treatment conditions suggests transcriptional alterations are mediated by ER α /chromatin binding induced changes conferred by neoadjuvant tamoxifen treatment.

We present the first comprehensive assessment of DNA copy number, gene expression patterns and ER α /chromatin profiles at two different time points of breast cancer therapy: before and after neoadjuvant tamoxifen treatment. The patient series investigated in this work is relatively small, however utilizing the power of paired treatment samples we are able to determine molecular changes conferred by neoadjuvant tamoxifen treatment. Large-scale alterations of ER α action were observed due to therapy, resulting in a substantial synchronization of ER α /chromatin binding and gene expression patterns between

patients. We uncovered evidence these synchronized sites may be important in breast cancer outcome as they show more binding in metastatic samples *versus* primary tumor samples. In addition, we found FOXA1 binding at these sites to be dynamic in response to estradiol and tamoxifen. Binding profiles at these sites implicate EGF as a potential regulator of these sites. FOXA1 dynamic binding is linked to outcome and critically, we identify genes associated with tamoxifen-induced sites to be capable of differentiating patients for tamoxifen benefit.

With this, we illustrate that hormonal therapy in breast cancer reprograms the genomic behavior of the drug target, ER α , and consequently affects downstream proliferation gene programs linked to patient outcome. Due to the direct effects of therapeutics on transcription factor behavior and transcriptional output, biomarker discovery studies may be further facilitated by performing such studies after neoadjuvant drug exposure.

MATERIALS AND METHODS

Patients and characteristics

ER α -positive breast cancer patients were recruited as part of the ongoing AFTER study (Anastrozole, Fulvestrant or Tamoxifen Exposure Response in molecular profile, ClinicalTrials.gov #NCT00738777) at two Dutch hospitals (Netherlands Cancer Institute-Antoni van Leeuwenhoek (NKI-AVL; Amsterdam) and Radboud University Medical Center (RadboudUMC; Nijmegen). Patient accrual occurred between August 2008 and February 2013. Local medical ethical authorities at both centers approved of the collection protocols. Pre-menopausal and male patients were treated with tamoxifen. Post-menopausal patients were assigned randomly to tamoxifen, anastrozole or fulvestrant therapy. Among them, only data from tamoxifen treated patients were analyzed (Supplemental Figure 1B). Patients were eligible if they had invasive, non-inflammatory breast cancer and were treated with hormonal therapy if the tumor was hormone receptor-positive by immunohistochemical (IHC) staining at diagnosis. Patients were excluded if they had multi-centric or metastatic breast cancer or if they received hormone replacement therapy in the previous 12 months. Patients were treated during the time between core-needle biopsy for diagnosis and surgery (typically 2-6 weeks). In a normal setting tamoxifen can take up to 8 weeks to reach therapeutic, steady-state levels in the blood plasma [21, 22]. Based on pharmacokinetics studies by Fabian *et al*, we chose a loading dose of 40mg orally, twice daily during 7 days with follow-up standard dosage of 20mg orally once daily to be able to reach steady state levels within 2 weeks [21].

We attempted to collect both pre-treatment and post-

treatment fresh-frozen (FF) and formalin-fixed, paraffin-embedded (FFPE) material for each patient (Supplemental Figure 1A). Sample fixation for each assay is listed in the methods details below. Ten serial sections of 5 μ m each were taken from both the core-needle biopsy (14 gauge) and primary tumor FFPE material for analyses. Pre-treatment FF material was in the form of core-needle biopsy made up of 10-15 serial sections of 30 μ m each. Post-treatment FF material was sectioned in 30 serial sections of 30 μ m each from the primary tumor taken at the time of surgery. All replicate experiments are from additional sections taken from the same tumor.

Immunohistochemical staining and assessment

Immunohistochemistry of samples was performed on a BenchMark Ultra autostainer (Ventana Medical Systems) for ER α , Progesterone Receptor (PR) and Receptor Tyrosine-Protein Kinase erbB2 (HER2). Briefly, paraffin sections were cut at 3 μ m, heated at 75°C for 28 minutes and deparaffinized in the instrument with EZ prep solution (Ventana Medical Systems). Heat-induced antigen retrieval was carried out using Cell Conditioning 1 (CC1, Ventana Medical Systems) for 36 minutes at 95°C (ER α , PR and HER2) or 32 minutes at 95°C (Ki-67). ER α was detected using clone SP1 (ready-to-use dispenser, 32 minutes room temperature (RT), Roche), PR with clone 1E2 (ready-to-use dispenser, 32 minutes RT, Roche), HER2 with clone 4B5 (ready-to-use dispenser, 12 minutes RT, Roche) and Ki67 detection using clone MIB1 (1:250 dilution, 32 minutes RT, DAKO). Bound antibody (ER α , PR and HER2) was detected using the UltraView Universal DAB Detection Kit (Ventana Medical Systems), while detection for Ki67 was visualized using the OptiView DAB Detection Kit (Ventana Medical Systems). Slides were counterstained with hematoxylin. ER α /PR/HER2/Ki67 scoring was performed on whole slides by a single pathologist (JW) blinded to patient status. For ER α , PR and Ki67 the percent of positive tumor nuclei was determined. For ER α /PR, 10% was used as a cut-off for positive according to current standard European guidelines. HER2 scoring was performed on a 0 to 3+ intensity scale examined in the nuclei (0, negative staining; 1+, weak staining; 2+ moderate staining, 3+, strong staining). There were no \geq 2+ HER2 samples in the cohort.

Chromatin immunoprecipitation and sequencing

Chromatin immunoprecipitations (ChIP) were carried out on FF material as previously described [23] using 5 μ g of ER α antibody (SC-543; SantaCruz Biotechnology) and 50 μ l Dynabeads. Subsequently, Illumina-indexed libraries were constructed for each sample using the ER α ChIP DNA (ChIP-seq) (Figure 2A).

Indexed DNA from 9 11 samples was equimolarly pooled sequenced on a single Illumina HiSeq2000 flowcell lane. Single-end 50bp reads were generated for each sample. Raw sequence data were aligned to the human genome (Ensembl 37) using Burrows-Wheeler Aligner (BWA) (mapping quality \geq 20, duplicate reads removed). As control, input chromatin was used. For each treatment category, 6-10 control samples were pooled equimolarly together to create a meta-pool of control input material. This meta-pool material was used as a reference sample for further analysis.

Analysis of ChIP-sequencing

To identify enriched genomic regions we used two peak callers, MACS [24] and DFilter [25]. We ran MACS using the default parameters with the exception of a more significant *p*-value threshold at 1.00e-7. DFilter settings were default for transcription factor detection. The intersection of the peaks from both peak callers was used for the final list of enriched regions (peaks) for each sample. Treatment conditions were analyzed separately for peaks where the corresponding treatment control input was used as reference.

We used the R Bioconductor 'DiffBind' package [26] to generate Venn diagrams of peaks for samples to determine a core set of binding events (identified in at least 2 of the samples in the set). This core set of binding events was used to identify the top DNA binding motifs using SeqPos from the Galaxy Cistrome package (<http://cistrome.org/ap/>) with z-score threshold set at -3.09, which corresponds to a *p*-value of 0.001 [27]. Other binding events were analyzed for DNA binding motifs using SeqPos from the Galaxy Cistrome package in the same manner. In addition, CEAS analysis was used from the Galaxy Cistrome package to characterize the genomic regions at binding sites.

To determine the intra-sample reproducibility of the ChIP-seq data we performed ER α ChIP-seq replicates on the same tumor sample using the same conditions. Additional data were analyzed for 3 supplementary ER α positive samples (not in the AFTER study) of similar overall read number. Read counts in 2,262 known non-overlapping ER α binding regions [26] were determined from the ChIP-seq alignment files using BEDTools coverageBed [28]. A correlation matrix of the read count data for all 5 samples (2 replicate pairs, 3 un-related) was then generated using the 'cor' function in R version 3.1.2. To determine the alternate allele frequency of variants in ChIP-seq data variants were called in ChIP-seq data of 4 treatment pairs using BWA, BCFTTools and SAMTools. Only heterozygous single nucleotide variants with \geq 10 alternate reads and \geq 10 reference reads were chosen for further analysis. Alternate allele frequency is defined as the DP4 (SAMTools) alternate allele count divided by the DP4 total reads count at each SNV.

For analysis of public ChIP-seq datasets, raw sequence data were downloaded from GEO (GSE25315, GSE25316, GSE32222, GSE68355) and aligned and MQ filtered as above. We used SeqMINER version 1.3.3 [29] to visualize the aligned data at regions of interest (bed format).

DNA copy number profiling

We employed the CopywriteR method [30] to detect DNA copy number from ‘off-target’ sequence reads taken from the ChIP-seq data without a reference sample. Circular binary segmentation was used to segment the data as previously described [30, 31].

RNA isolations, microarray hybridization and analysis

Total RNA was isolated from formalin fixed paraffin embedded tissue as described previously [32]. After DNase treatment samples were purified using the Qiagen RNeasy FFPE kit. Total RNA (50ng) was reversed transcribed, amplified (Rubicon; C-WTA kit C) labeled with Cy3 (Genomic DNA enzymatic Labeling kit; Agilent Technologies) and subsequently purified (Amicon ultra 30kDa filters). Cy3-labeled cDNA was hybridized to custom full genome arrays—array design based on Agilent Catalog #G2514F—at 65°C for 17 hours and subsequently washed. Arrays were scanned with a dual laser scanner (Agilent Technologies). Image analysis of the scanned arrays was performed to quantify fluorescent intensities using Feature Extraction software version 9.5 (Agilent Technologies).

Feature signal intensities were processed, imputation of missing values and summarization of all genes with multiple probes was performed as previously described [33]. Visualization of the data was performed with Partek Genomics Suite (Partek) using Hierarchical Clustering (Pearson Dissimilarity, average linkage) with centered data. Top variable genes were selected based on variance across samples (> 1). An ANOVA analysis was used to determine differentially expressed genes between treatment conditions of the top variable genes (Partek). Significant genes were selected univariately with FDR < 0.001 and the log₂ space with fold change > 2 .

To examine the relationship between published proliferation gene expression module scores and MKI67 gene expression levels we calculated the AURKA and CIN70 proliferation gene expression module scores (containing proliferation-associated genes (PAGs)) as described in the work by Martin and Dowsett and colleagues [34]. The percent change as described previously [34] between pre- and post-treatment gene expression module scores and MKI67 gene expression values were calculated and visualized with a scatterplot.

In addition, MammaPrint® scores were calculated [35].

DNA copy number and gene expression integration

We used the R package ‘iC10’ to implement the classifier, IntClust to integrate and classify the samples into the 10 IntClust groups using DNA copy number and gene expression data [36, 37]. Classification was carried out as described previously using default settings [36].

Statistical analysis

All statistical analyses were carried out in R version 3.1.2 (<http://www.R-project.org>) including all patients (pre-/post-menopausal and male) unless otherwise noted. Strength of associations between continuous variables was calculated with the Wilcoxon test, either Wilcoxon-rank sum (paired data) or Wilcoxon signed rank (non-paired data). For statistical analysis of categorical IntClust data, Pearson’s chi-square test was used with simulated *p*-value based on 2000 replicates. For examining the uncertainty in DNA copy number hierarchical analysis we used the R package ‘pvclust’ (distance = correlation, clustering = Ward; number of iterations = 1000). Briefly, *p*-values were calculated *via* two methods, AU (approximately unbiased) and BP (Bootstrap probability). The AU method computes using multiscale bootstrap resampling and the BP method computes by normal bootstrap resampling. The resultant clusters were depicted visually as dendrograms with AU and BP values shown above the branch. To examine the relationship between the percent change in MKI67 gene expression and gene expression module percent change Spearman’s *rho* was calculated. In addition, ChIP-seq sample replicate correlation using read counts in known ERα binding regions [26] was examined by Spearman’s *rho*.

For survival analysis, normalized gene expression datasets were downloaded from GEO: GSE6532 [38], GSE22219 [39], GSE2034 [40] and GSE1121 [41]. From the tamoxifen treated datasets [38, 39], ERα positive, tamoxifen treated patients were selected, respectively ($n = 250$, $n = 134$). From the non-endocrine datasets [40, 41], ERα positive untreated patients were selected, respectively ($n = 209$, $n = 158$). ERα status for the Schmidt *et al.* dataset was determined as described previously [42]. For each dataset, we selected all probes matching the 96 genes associated with 126 tamoxifen-synchronized sites (disregarding genes not present on the array). The patients were then stratified into two groups with the selected gene expression data using unsupervised hierarchical Ward linkage clustering with Pearson correlation distance. Following that we calculated the hazard ratio to determine significant difference in distant metastasis-free survival (DMFS) between the two groups and Kaplan Meier graphs

were constructed to visualize the data.

ACKNOWLEDGMENTS

We thank all patients who contributed to this study. We also thank all the hospital staff at both centers, RadboudUMC and the Netherlands Cancer Institute (NKI). The authors would also like to acknowledge the effort and support of the NKI Genomics Core Facility and the NKI Core Facility Molecular Pathology & Biobanking (CFMPB) for supplying NKI Biobank material and/or lab support.

GRANT SUPPORT

This work was supported by A Sister's Hope; and partially funded by a TI-Pharma Value Creation grant [T3-502]. Funding for open access charge: A Sister's Hope.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

Editorial note

This paper has been accepted based in part on peer-review conducted by another journal and the authors' response and revisions as well as expedited peer-review in *Oncotarget*.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer J Int Cancer*. 2015 Mar 1;136:E359-86.
2. Fox MS. On the diagnosis and treatment of breast cancer. *JAMA J Am Med Assoc*. 1979 Feb 2;241:489-94.
3. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A, Trent J, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med*. 2001 Feb 22;344:539-48.
4. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, et al. Molecular portraits of human breast tumours. *Nature*. 2000 Aug 17;406:747-52.
5. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001 Sep 11;98:10869-74.
6. Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer*. 2007 Oct;7(10):791-9.
7. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002 Jan 31;415:530-6.
8. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002 Dec 19;347:1999-2009.
9. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, Nevins JR. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A*. 2001 Sep 25;98:11462-7.
10. Colleoni M, Rotmensz N, Maisonneuve P, Mastropasqua MG, Luini A, Veronesi P, Intra M, Montagna E, Cancellato G, Cardillo A, Mazza M, Perri G, Iorfida M, et al. Outcome of special types of luminal breast cancer. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 2012 Jun;23:1428-36.
11. Droog M, Beelen K, Linn S, Zwart W. Tamoxifen resistance: from bench to bedside. *Eur J Pharmacol*. 2013 Oct 5;717:47-57.
12. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res BCR*. 2007;9:R6.
13. Ignatiadis M, Sotiriou C. Luminal breast cancer: from biology to treatment. *Nat Rev Clin Oncol*. 2013 Sep;10:494-506.
14. Orlando L, Schiavone P, Fedele P, Calvani N, Nacci A, Rizzo P, Marino A, D'Amico M, Sponziello F, Mazzoni E, Cinefra M, Fazio N, Maiello E, et al. Molecularly targeted endocrine therapies for breast cancer. *Cancer Treat Rev*. 2010 Nov;36 Suppl 3:S67-71.
15. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*. 1998 Dec 23;95:927-37.
16. Osborne CK. Tamoxifen in the treatment of breast cancer. *N Engl J Med*. 1998 Nov 26;339:1609-18.
17. Amaral C, Varela C, Borges M, Tavares da Silva E, Roleira FMF, Correia-da-Silva G, Teixeira N. Steroidal aromatase inhibitors inhibit growth of hormone-dependent breast cancer cells by inducing cell cycle arrest and apoptosis. *Apoptosis Int J Program Cell Death*. 2013 Nov;14:26-36.
18. Thürlimann B, Robertson JFR, Nabholz JM, Buzdar A, Bonnetterre J, Arimidex Study Group. Efficacy of tamoxifen following anastrozole ('Arimidex') compared with anastrozole following tamoxifen as first-line treatment

- for advanced breast cancer in postmenopausal women. *Eur J Cancer Oxf Engl*. 2003 Nov;39:2310-7.
19. Ellis MJ, Tao Y, Luo J, A'Hern R, Evans DB, Bhatnagar AS, Chaudri Ross HA, von Kameke A, Miller WR, Smith I, Eiermann W, Dowsett M. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *J Natl Cancer Inst*. 2008 Oct 1;100:1380-8.
 20. Goncalves R, Ma C, Luo J, Suman V, Ellis MJ. Use of neoadjuvant data to design adjuvant endocrine therapy trials for breast cancer. *Nat Rev Clin Oncol*. 2012 Apr;9:223-9.
 21. Fabian C, Sternson L, Barnett M. Clinical pharmacology of tamoxifen in patients with breast cancer: comparison of traditional and loading dose schedules. *Cancer Treat Rep*. 1980 Jul;64:765-73.
 22. Spooner D, Fairlamb D, Fermont D. A multicentre trial to compare a loading dose regimen of tamoxifen ('Nolvadex') with conventional dosing in the treatment of advanced breast cancer. The British Tamoxifen Study Group. *Clin Oncol R Coll Radiol G B*. 1991 May;3:133-6.
 23. Zwart W, Koornstra R, Wesseling J, Rutgers E, Linn S, Carroll JS. A carrier-assisted ChIP-seq method for estrogen receptor-chromatin interactions from breast cancer core needle biopsy samples. *BMC Genomics*. 2013;14:232.
 24. 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.
 25. Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, Prabhakar S. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat Biotechnol*. 2013 Jul;31:615-22.
 26. 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, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012 Jan 19;481:389-93.
 27. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y, Pape UJ, Poidinger M, Chen Y, et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol*. 2011;12(8):R83.
 28. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinforma Ed Board Andreas Baxeavanis Al*. 2014;47:11.12.1-11.12.34.
 29. Ye T, Krebs AR, Choukrallah M-A, Keime C, Plewniak F, Davidson I, Tora L. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res*. 2011 Mar;39:e35.
 30. Kuilman T, Velds A, Kemper K, Ranzani M, Bombardelli L, Hoogstraat M, Nevedomskaya E, Xu G, de Ruyter J, Lolkema MP, Ylstra B, Jonkers J, Rottenberg S, et al. CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biol*. 2015;16:49.
 31. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostat Oxf Engl*. 2004 Oct;5:557-72.
 32. Sapino A, Roepman P, Linn SC, Snel MHJ, Delahaye LJMJ, van den Akker J, Glas AM, Simon IM, Barth N, de Snoo FA, van 't Veer LJ, Molinaro L, Berns EMJJ, et al. MammaPrint molecular diagnostics on formalin-fixed, paraffin-embedded tissue. *J Mol Diagn JMD*. 2014 Mar;16:190-7.
 33. Severson TM, Peeters J, Majewski I, Michaut M, Bosma A, Schouten PC, Chin S-F, Pereira B, Goldgraben MA, Bismeyer T, Kluin RJC, Muris JFF, Jirstrom K, et al. BRCA1-like signature in triple negative breast cancer: Molecular and clinical characterization reveals subgroups with therapeutic potential. *Mol Oncol*. 2015 May 7;
 34. Gao Q, Patani N, Dunbier AK, Ghazoui Z, Zvelebil M, Martin L-A, Dowsett M. Effect of aromatase inhibition on functional gene modules in estrogen receptor-positive breast cancer and their relationship with antiproliferative response. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2014 May 1;20:2485-94.
 35. Glas AM, Floore A, Delahaye LJMJ, Witteveen AT, Pover RCF, Bakx N, Lahti-Domenici JST, Bruinsma TJ, Warmoes MO, Bernards R, Wessels LFA, Van't Veer LJ. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics*. 2006;7:278.
 36. Ali HR, Rueda OM, Chin S-F, Curtis C, Dunning MJ, Aparicio SA, Caldas C. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol*. 2014;15:431.
 37. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012 Jun 21;486:346-52.
 38. Loi S, Haibe-Kains B, Desmedt C, Lallemant F, Tutt AM, Gillet C, Ellis P, Harris A, Bergh J, Foekens JA, Klijn JGM, Larsimont D, Buyse M, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol Off J Am Soc Clin Oncol*. 2007 Apr 1;25:1239-46.
 39. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res*. 2011 Sep 1;71:5635-45.
 40. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EMJJ, Atkins D, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005 Feb 19;365:671-9.
 41. Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, Lehr H-A, Hengstler JG, Kölbl H, Gehrman M. The humoral immune system has a key prognostic impact

- in node-negative breast cancer. *Cancer Res.* 2008 Jul 1;68:5405-13.
42. Oosterkamp HM, Hijmans EM, Brummelkamp TR, Canisius S, Wessels LFA, Zwart W, Bernards R. USP9X downregulation renders breast cancer cells resistant to tamoxifen. *Cancer Res.* 2014 Jul 15;74:3810-20.
 43. Polley M-YC, Leung SCY, McShane LM, Gao D, Hugh JC, Mastropasqua MG, Viale G, Zabaglo LA, Penault-Llorca F, Bartlett JMS, Gown AM, Symmans WF, Piper T, et al. An international Ki67 reproducibility study. *J Natl Cancer Inst.* 2013 Dec 18;105:1897-906.
 44. Jansen MPH, Knijnenburg T, Reijm EA, Simon I, Kerkhoven R, Droog M, Velds A, van Laere S, Dirix L, Alexi X, Foekens JA, Wessels L, Linn SC, et al. Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res.* 2013 Nov 15;73:6632-41.
 45. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoutte J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, et al. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet.* 2006 Nov;38:1289-97.
 46. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EGY, Huang PYH, Welboren W-J, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature.* 2009 Nov 5;462:58-64.
 47. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell.* 2005 Jul 15;122:33-43.
 48. Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, Zaret KS. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* 1998 Jan 2;17:244-54.
 49. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell.* 2002 Feb;9:279-89.
 50. 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.* 2011 Jan;43:27-33.
 51. Laganière J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguère V. From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci U S A.* 2005 Aug 16;102:11651-6.
 52. 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 Mar 21;132:958-70.
 53. Mohammed H, Russell IA, Stark R, Rueda OM, Hickey TE, Tarulli GA, Serandour AA, Serandour AAA, Birrell SN, Bruna A, Saadi A, Menon S, Hadfield J, et al. Progesterone receptor modulates ER α action in breast cancer. *Nature.* 2015 Jul 16;523:313-7.
 54. Franco HL, Nagari A, Kraus WL. TNF α signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Mol Cell.* 2015 Apr 2;58:21-34.
 55. Liu Z, Merkurjev D, Yang F, Li W, Oh S, Friedman JM, Song X, Zhang F, Ma Z, Ohgi KA, Kronen A and Rosenfeld MG. Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* 2014 Oct 9;159(2):358-73.
 56. Stelloo S, Nevedomskaya E, van der Poel HG, de Jong J, van Leenders GJ, Jenster G, Wessels LF, Bergman AM, Zwart W. Androgen receptor profiling predicts prostate cancer outcome. *EMBO Mol Med.* 2015 Sep 27;
 57. 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, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet.* 2013 Dec;45:1446-51.
 58. 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, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet.* 2013 Dec;45:1439-45.
 59. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012 Mar 8;366:883-92.
 60. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature.* 2012 Jun 21;486:395-9.
 61. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature.* 2014 Aug 14;512:155-60.
 62. Osborne CK, Schiff R. Mechanisms of endocrine resistance in breast cancer. *Annu Rev Med.* 2011;62:233-47.
 63. Polychronis A, Sinnott HD, Hadjiminas D, Singhal H, Mansi JL, Shivapatham D, Shousha S, Jiang J, Peston D, Barrett N, Vigushin D, Morrison K, Beresford E, et al. Preoperative gefitinib versus gefitinib and anastrozole in postmenopausal patients with oestrogen-receptor positive and epidermal-growth-factor-receptor-positive primary breast cancer: a double-blind placebo-controlled phase II randomised trial. *Lancet Oncol.* 2005 Jun;6:383-91.
 64. Osborne CK, Neven P, Dirix LY, Mackey JR, Robert J, Underhill C, Schiff R, Gutierrez C, Migliaccio I,

Anagnostou VK, Rimm DL, Magill P, Sellers M. Gefitinib or placebo in combination with tamoxifen in patients with hormone receptor-positive metastatic breast cancer: a randomized phase II study. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2011 Mar 1;17:1147-59.

65. Cristofanilli M, Valero V, Mangalik A, Royce M, Rabinowitz I, Arena FP, Kroener JF, Curcio E, Watkins C, Bacus S, Cora EM, Anderson E, Magill PJ. Phase II, randomized trial to compare anastrozole combined with gefitinib or placebo in postmenopausal women with hormone receptor-positive metastatic breast cancer. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2010 Mar 15;16:1904-14.