Acquired CYP19A1 amplification is an early specific mechanism of aromatase inhibitor resistance in ERα metastatic breast cancer

Luca Magnani[#,*,1], **Gianmaria Frige**[#2], **Raffaella Maria Gadaleta**[1], **Giacomo Corleone**[1], **Sonia Fabris**[3], **Mannus H. Kempe**[4], **Pernette J. Vershure**[4], **Iros Barozzi**[5], **Valentina Vircillo**[6], **Sung-Pil Hong**[1], **Ylenia Perone**[1], **Massimo Saini**[7], **Andreas Trumpp**[7], **Giuseppe Viale**[8], **Antonino Neri**[3,9], **Simak Ali**[1], **Marco Angelo Colleoni**[10], **Giancarlo Pruneri**[8,*], and **Saverio Minucci**[2,11,*]

[1]Department of Surgery and Cancer, Imperial College London, London, UK

[2]Department of Experimental Oncology, European Institute of Oncology Milan, Italy

[3]Hematology Unit, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, IT

[4]Swammerdam Institute for Life Sciences, University of Amsterdam

[5]Genomics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA

[6]Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Arcavacata di Rende (CS), Italy

[7]Division of Stem Cells and Cancer, Deutsches Krebsforschungszentrum (DKFZ), and Institute for Stem Cell Technology and Experimental Medicine GmbH, Heidelberg, Germany

[8]Division of Pathology, European Institute of Oncology and University of Milan, School of Medicine, Milan, Italy

[9]Department of Oncology and Hemato-oncology, University of Milano, Milan, IT

[10]Division of Medical Senology, European Institute of Oncology, IEO, Milan, Italy

[11]Department of Biosciences, University of Milano, Milan, IT

[#] These authors contributed equally to this work.

Abstract

[*]To whom correspondence should be addressed: saverio.minucci@ieo.eu, giancarlo.pruneri@ieo.it and l.magnani@imperial.ac.uk.

Tumor evolution is shaped by many variables, potentially involving external selective pressures induced by therapies[1]. After surgery, estrogen receptor (ERα) positive breast cancer (BCa) patients are treated with adjuvant endocrine therapy[2] including selective estrogen receptor modulators (SERMs) and/or aromatase inhibitors (AIs)[3]. However, over 20% of patients relapse within 10 years and eventually progress to incurable metastatic disease[4]. Here we demonstrate that the choice of therapy has a fundamental influence on the genetic landscape of relapsed diseases: in this study, 21.5% of AI-treated, relapsed patients had acquired *CYP19A1* gene (aromatase) amplification (*CYP19A1^{amp}*). Relapsed patients also developed numerous mutations targeting key breast cancer genes including *ESR1* and *CYP19A1*. Strikingly, *CYP19A1^{amp}* cells also emerge *in vitro* but only in AI resistant models. *CYP19A1* amplification causes increased aromatase activity and estrogen-independent ERα binding to target genes resulting in *CYP19A1^{amp}* cells displaying decreased sensitivity to AI treatment. Collectively these data suggest that AI treatment itself selects for acquired *CYP19A1* amplification and promotes local autocrine estrogen signalling in AI resistant metastatic patients.

ERα activation characterizes over 70% of BCa where it represents the key prognostic factor and therapeutic target[5]. ERα activation is primarily dependent on circulating estrogens and results in genome-wide chromatin binding at thousands of regulatory regions[6]. ERα binding leads to the transcription of hundreds of genes central to BCa growth[6]. Endocrine therapies including SERMs and AIs were developed to prevent ERα activation and block BCa growth[5]. The mechanisms behind drug resistance are only partially understood and often involve transcriptional activation of alternative survival pathways, especially at later stages of the disease[7]. Nonetheless, recent genomic studies highlight how ERα signalling might still play a role in metastatic disease. For example, activating somatic mutations targeting *ESR1* (the gene encoding ERα) are found at higher frequencies after endocrine therapy[8,9]. These mutations have been characterized in metastatic lesions from patients that received several cycles of ET and chemotherapy[10,11], suggesting that the selective pressure imposed by endocrine treatments might favour the development of focused genetic aberrations during tumour evolution[11]. It is however impossible to infer from most studies when genetic aberrations originate and how these are selected, since patients are biopsied after multiple treatments. While the SERM Tamoxifen (TAM) directly blocks ERα co-activation in the tumor cell, AI targets CYP19A1 (aromatase) in the peripheral tissue thereby lowering estrogen availability. We recently reported that ERα positive BCa cells activate alternative epigenetic programs in response to TAM or AI[12] suggesting that choice of endocrine therapies might contribute to tumor evolution. Here we examine, in parallel and for the first time, a cohort of estrogen receptor positive α patients who were treated with single agent adjuvant endocrine therapies (either TAM or non-steroidal AI) and re-biopsied whenever they had their first distal relapse (Fig 1A and Supplementary Figures S1-2).

We initially assessed copy number alterations (CNAs) of the genes encoding the targets of AI and TAM *(CYP19A1:*15q21 and *ESR1*:6q25 respectively) considering the central role of copy number changes in breast cancer[13]. Meta-analysis of previously published data from primary, treatment-naive patients using GISTIC-based[14] cBioPortal[15] shows that *CYP19A1* and *ESR1* CNAs are exceedingly rare in ERα positive primary BCa (0.006%, 2/321 for *CYP19A1* and 0.018%, *6/321* for *ESR1* in ERα positive primary BCa, The Cancer Genome

Atlas (TCGA) CNAs data 16, threshold: 1.5 fold change). Using an independent database of SNP-array based studies with an alternative CNAs algorithm17 confirms the rarity of *CYP19A1* amplification events (Supplementary table 1). *CYP19A1* and *ESR1* amplification are also rare in other primary cancers (Supplementary Figures S3A-B and Supplementary table 1). These data demonstrate that *CYP19A1* and *ESR1* loci are not re-arrangement hotspots in untreated primary cancers. We then analyzed our discovery cohort consisting of tumor samples collected from the first relapse after single therapy using a TaqMan CNA assay comparing metastatic with matched normal breast tissue. Strikingly, we find that the *CYP19A1* locus is amplified (*CYP19A1amp*) in 6/37 (16%) of the patients that received AI (*CYP19A1amp*). Conversely, only one patient (3%) that received TAM has evidence for *CYP19A1* amplification (Fig. 1A). The *ESR1* locus is also significantly amplified in relapsed material (24% and 13%, AI and TAM-treated cohorts respectively, Fig. 1A). To confirm these data, we then investigated an independent validation cohort with similar clinical characteristics. In agreement with the discovery cohort, we find that *CYP19A1* is amplified in 6/19 (32%) of AI treated patients and only 1/19 (5%) of TAM-treated patients (Fig. 1B). *ESR1* is amplified in 4/19 (21%) of AI treated and 0/19 of the Tamoxifen-treated relapses (Supplementary Figure S4A). The *CYP19A1* locus shows evidence for both focal and chromosome-wide amplification (Supplementary Figure S5A). *CYP19A1* and *ESR1* CNAs might work cooperatively considering the rate of co-amplification in AI treated patients (8/12 *CYP19A1amp* patients also carry *ESR1amp*, Supplementary Figure S5B). Notably, we could identify *CYP19A1* and *ESR1* amplification also in patient-derived xenografts (PDXs) obtained from patients previously treated with non-steroidal AI (Fig. 1C and Supplementary Figure S4B). Collectively these data show that treatment with reversible AI significantly increase the frequency of *CYP19A1* amplification at first distal relapse (21.5% vs 4%, AI vs. Tamoxifen, P= 0.009, P=0.004 including PDXs, two-tailed Fisher's Exact test). Similarly, we observe a trend for AI treated patients to preferentially amplify the *ESR1* locus (23% vs 8%, AI vs. Tamoxifen P= 0.06, P=0.03 including PDXs, two-tailed Fisher's Exact test). *CYP19A/ESR1* amplification in distal relapses from AI resistant BCa is strongly reminiscent of Androgen Receptor amplification in castration resistant prostate cancer patients18,19.

We next designed a DNA-FISH assay to validate *CYP19A1* amplification, and to investigate its degree of heterogeneity. We examined 4 cases found to be amplified by TaqMan: all of them present strong evidence for cluster amplification (Fig. 2A-B). FISH analysis also confirm 100% of TaqMan calls in the validation dataset (Supplementary figure S6). Over 90% of nuclei from each of the metastatic samples examined by FISH has *CYP19A1* amplification signals, indicating that *CYP19A1amp* cells represent the dominant clone. Additionally, the *15α Satellite/CYP19A1* ratio strongly suggests that *CYP19A1* amplification is not a consequence of unspecific aneuploidy (Fig. 2B). Using DNA-FISH we could not find convincing evidence of *CYP19A1* amplification in the respective primary samples (Fig. 2A, see online methods). Therefore, these results support the notion that *CYP19A1* amplification occurs under treatment, although we cannot exclude the presence of very small *CYP19A1amp* subclones at diagnosis.

We then investigated the frequency of metastatic-specific ERα activating mutations11 and other commonly occurring mutations13,20 (at *PIK3CA, MAPK, TP53 and GATA3* genes) in

AI or TAM treated distal relapses compared to matched normal germline DNA using an AmpliSeq Custom panel for targeted sequencing (Supplementary Table 2, see online methods). Overall, we found similar patterns of mutations between the two cohorts (Fig. 2C and Extended Data 1-2). These patterns however are potentially different from the ones previously characterized in primary BCa (Supplementary Figure S7). For example, we could identify several novel ERα mutations including the predicted activating mutations pLeu536His21 (TAM 19T), pMet543Ala22 (AI 28T and AI 46T) pAsp538Ala (TAM 10T) and p.Arg503Gln (AI 28T and AI 39T) (Fig 2C-D and Extended Datasets 1-2) in addition to the common p.Tyr537Ser (AI 36T). These mutations occur at relatively higher allele frequencies and can be polyclonal (AI 28T) (Fig S8A). In addition, we could identify several truncating TP53 mutations (examples: p.Arg213*, AI 8T; p.Cys242*, AI 13T and p.Glu294* TAM 23T; Fig 2C-D and Extended Datasets 1-2). Most of these mutations were confirmed by a second assay (Supplementary figure S8B). Interestingly, we designed probes against CYP19A1 and identified the novel recurrent mutation pPro410Ser/Leu (AI 5T and 18T, Supplementary figure S8B and Extended Datasets 2). This mutation is mutually exclusive with *CYP19A1* amplification further suggesting its functional significance (Supplementary Figure S8B). Interestingly, p.Pro410 might be involved in the formation of a functional channel23. Collectively, these data demonstrate that several metastatic-specific mutations might play a role early on during tumor progression.

In post-menopausal women, AI target the peripheral conversion of testosterone/ androstenedione to estrogens by inhibiting the product of the *CYP19A1* gene, the aromatase enzyme. Cholesterol is the common precursor of sex hormones including testosterone and androstenedione, the two substrates of the aromatase enzyme24. *In vitro,* AI resistant cells develop *de novo* cholesterol biosynthesis via epigenetic events to promote autonomous ERα activation12. Interestingly, current treatment protocols can effectively remove all estradiol from serum used in culture, while they are far less efficient in removing testosterone25. Thus, we hypothesized that *CYP19A1* amplification might favour autocrine estrogens production starting from epigenetically driven *de-novo* cholesterol biosynthesis or residual male hormones. The possibility that there might be some CNAs contributing to AI resistance was investigated using a shallow whole genome sequencing approach26 to profile several cell lines derived from ERα positive BCa MCF7 cells and exposed to distinct endocrine treatments12,27 (Supplementary Figure S9A). Among those, LTED cells are derived upon chronic estrogen deprivation27,28, mimicking AI treated BCa. Genomic analysis suggests that all endocrine resistant cells maintain DNA re-arrangements of parental MCF7 cells (Chr3p-20q translocation). However, LTED cells also acquired CNAs around the *CYP19A1* locus (Supplementary Figure S9B). *CYP19A1* CNAs were exclusively identified in LTED but not in Tamoxifen-Fulvestrant resistant models (Supplementary Figure S9B). We next quantified *CYP19A1* CNAs using our TaqMan assay (Fig. 3A) and DNA-FISH (Fig. 3B). Remarkably, the level of amplification was comparable to the one developed by AI-treated patients *in vivo* (Fig. 1C and 3A). Of note, we could not identify a single *CYP19A1*^{amp} cell in MCF7 cells, suggesting that *CYP19A1* CNAs were acquired during treatment. LTED cells also undergo a tenfold increase in ERα protein levels12 although we could not find significant evidence for *ESR1* amplification or ERα activating mutations. Collectively, these data demonstrate that estrogen deprivation promotes *CYP19A1* CNAs *in vitro*, in a way

similar to AI treatment *in vivo*. *CYP19A1^{amp}* LTED cells have higher levels of both CYP19A1 mRNA and protein compared to *CYP19A1^{wt}* MCF7 cells (Fig. 3C-E). Single-cell RNA-FISH (see online protocol) demonstrates a large degree of heterogeneity in CYP19A1 expression, however *CYP19A1^{amp}* cells have a significantly higher chance of accumulating more CYP19A1 mRNA molecules per cell (Fig. 3E). We also found that LTED cells have significantly higher aromatase activity when compared to MCF7 cells, and this can be partially antagonized using AI (Supplementary Figure S10A). *CYP19A1^{amp}* LTED cells, but not *CYP19A1^{wt}* MCF7 cells, display significant transcriptional activation of three well characterized estrogen target genes (e.g. TFF1, EGR3 and CA127) in response to androstenedione (Fig. 3F). More importantly, transcriptional activation by androstenedione can be partially blocked by AI (letrozole) (Fig. 3F). Collectively, these data indicate that acquired *CYP19A1* amplification drives increased endogenous aromatase activity and is required to promote estrogen independent transcription.

Consistently with transcriptional data, when we investigated the genome-wide ERα binding to target genes in MCF7 and LTED cells in the absence of estrogens, we found that the chromatin of LTED cells is strongly enriched in ERα binding compared to parental MCF7 cells (Fig. 4A). As expected, MCF7 cells grown in the presence of estrogen are not sensitive to AI (Fig. 4B) while *CYP19A1^{amp}* LTED cells have low sensitivity (IC50=80uM, Fig. 4C). To test the role of *CYP19A1^{amp}* on AI sensitivity, we then treated LTED cells and parental MCF7 cells with two independent siRNA targeting CYP19A1 and measured cell viability in response to the AI letrozole (Fig. 4B-D and Supplementary figure S10B). siCYP19A1 significantly increased sensitivity to AI treatment in *CYP19A1^{amp}* LTED cells (Fig. 4D) while did not affect MCF7 grown in estrogen-supplemented conditions. (Figure 4E and Supplementary Figure S10C). CYP19A1 over-expression did not confer any growth advantage to MCF7 cells grown in the presence of estradiol. However, CYP19A1 over-expression was sufficient to relieve cell cycle arrest in MCF7 cultured in absence of estrogens. Interestingly, this effect could not be antagonized by letrozole (Fig 4E). Finally, we confirmed that ERα still plays a role in *CYP19A1^{amp}* LTED cells growth as shown LTED sensitivity to Fulvestrant treatment (Supplementary Figure S10D). More importantly, treating *CYP19A1^{amp}* LTED cells with an irreversible AI (Exemestane) suggest that increased level of aromatase activity can be antagonized with a steroidal AI (Fig. 4F). Collectively, these data support our initial hypothesis and suggest that *CYP19A1* amplification might induce reduced sensitivity to reversible AI treatment.

CYP19A1 amplification triggers ERα activity by converting male sex hormones possibly obtained through endogenous epigenetic cholesterol biosynthesis12 or circulating within the tumor microenvironment. *In vivo* and *in vitro* data indicate that *CYP19A1* CNAs are *acquired* rather than selected. Indeed, LTED cells develop *CYP19A1* CNAs over the course of chronic estrogen deprivation (> 1 year). We then analyzed the time to first relapse merging our AI patient datasets. Interestingly, relapses characterized by *CYP19A1* CNAs emerged significantly later compared to WT *CYP19A1* (median 57 vs 30 months Amp vs. WT, p=0.0112 Log-Rank Mantel-Cox test). These data support the notion that breast cancer cells treated with AI slowly evolve to create a favourable autocrine microenvironment for themselves through genetic and epigenetic reprogramming (Fig. 4G). One unexplained clinical observation is that patients progressing under reversible AI treatment (letrozole/

anastrozole) occasionally respond to irreversible AI (i.e. exemestane). Interestingly, the disease of five *CYP19A1^{amp}*-AI patients within our study was stabilised for ~1-year on average after switching to exemestane (Supplementary Figures S1-2). Thus, it is tempting to speculate that *CYP19A1* amplification might arise in response to reversible inhibitors but could be antagonized by switching to irreversible inhibitors. Alternatively, it should be clinically feasible to antagonize directly the low levels of circulating male hormones commonly found in post-menopausal women. Considering that AI normally target peripheral tissues, our data also warrant for AI pharmacodynamics studies to evaluate the ability of this class of drugs to target directly tumor cells. Taken together, our clinical data demonstrate that the evolution of breast cancer is shaped by clinical intervention and thus advocate the development of treatment- and setting-specific biomarkers.

Data Availability

ChIP-seq data for ERα can be accessed under GEO profile GSE60517. Primary cancer datasets analyzed for CNAs in this manuscript were retrieved using http://www.cbioportal.org/. For breast cancer specific mutational analyses, we have used all available datasets from http://www.cbioportal.org/. Each study has been labelled according to publisher's instruction.

On line methods

Adjuvant setting patient selection and tissue preparation

The IEO Data Quality Control Unit selected from the Institutional database a set of consecutive breast cancer patients fulfilling the following criteria: i) tumors classified as Luminal A-like and Luminal B-like (HER2 negative), in accordance with St. Gallen 2013 recommendations 29; ii) patients receiving exclusively AIs (#50 pts.) or TAM (#50 pts.) as systemic adjuvant therapy; iii) patients with at least one year of follow-up; iv) patients who experienced a distant metastasis as first event after surgery and upon adjuvant therapy. Patients who initially presented with bilateral breast tumor, receiving neoadjuvant treatments, and with metastatic breast disease at the time of presentation or within 12 months after surgery have been excluded. The initial data set used for this project comprised 26,495 women who had undergone surgery for a first primary breast cancer at the IEO between 1994 and 2014. All the cases prospectively entered the IEO breast cancer database and were discussed at the weekly multidisciplinary meeting. Patients have been then followed up with physical examination every six months, annual mammography and breast ultrasound, blood tests every 6-12 months and further evaluations when symptomatic. Data on the patients' medical history, concurrent diseases, surgery, pathological evaluation, results of staging procedures, radiotherapy, adjuvant systemic treatments, events occurring during the follow-up and treatments for metastatic disease were available. Use of patients' data was approved by the ethics committee of the IEO and by the Italian Data Protection Authority. All patients signed an informed consent. All the primary tumors were fresh sampled, fixed in 4% buffered formalin and embedded in paraffin. All the metastatic biopsies were fixed in 4% buffered formalin. Detailed information regarding tumor type and grade, ER/PgR and Her-2 status and Ki-67 labelling index were available in all the cases of primary and

metastatic tumors. ER/PgR and HER2 immunoreactivity was assessed in line with the clinical practice procedures applicable at diagnosis. Her-2 immunoreactivity was assessed using the monoclonal antibody CB11 (Novocastra, 1:800) from 1995 till 2005, and the HercepTest (Dako) thereafter. Cases classified as Her-2 2+ by immunohistochemistry were tested by FISH analysis with Vysis probes, in accordance with the ASCO/CAP guidelines. Ki-67 labelling index was assessed by the Mib-1 monoclonal antibody (Dako, 1:200), by counting at least 500 invasive tumor cells, independent of their staining intensity and without focusing on hot-spots. Only tumors classified as Luminal A-like (ER and PgR positive, absence of Her-2 overexpression and Ki-67 <20%) and Luminal B-like (ER positive, Her-2 negative and at least one of Ki-67   20% and PgR <20%) in accordance with St. Gallen recommendations were included in the study29. All the samples (primary tumors and paired metastatic deposits) from the patients satisfying the aforementioned criteria have been reviewed at the IEO Division of Pathology (University of Milan) for assessing tumor cellularity and for tumor enrichment by macrodissection, if necessary. DNA from all the samples has been extracted by commercially available kits (QIAmp DNA FFPE Tissue Kit, QIAGEN) and the DNA yielding has been measured by Qubit Fluorimetric Quantitation (ThermoFisher Scientific). Finally, two 3 μm thick slides has been cut from all the samples and put on charged slides for FISH analysis.

DNA extraction and TaqMan assay

For tumor samples Hematoxylin/Eosin-stained sections were prepared assessing the percentage of tumor cells and evaluated by a pathologist. Samples with less than 80% of tumor cells were micro-dissected to increase the percentage of tumor cells. Matched normal DNA was extracted from from non-metastatic axillary lymph nodes or histologically non-neoplastic breast samples obtained from mammary quadrants macroscopically free of disease. Genomic DNA from FFPE tissue sections was extracted using QIAamp FFPE Tissue kit (Qiagen) according to the manufacturer's instructions. TaqMan Copy Number Assay (Applied Biosystems) for CYP19A1 gene (Hs00116110_cn) and ESR1 gene (Hs02488982_cn) was performed using the 7900HT Fast Real-Time PCR Systems (Applied Biosystems) according to the manufacturer's protocol. TERT, RNAsep, GARBR3 genes were analysed as endogenous reference genes (Applied Biosystems, cat# 4458373, cat# 4403326 and cat# hs_05365082_cn). Copy number for each sample was estimated by using the Copy Caller Software V1.0 (Applied Biosystems) using the matched normal counterpart as reference. CN range bars indicated the minimum and maximum CN calculated for the sample replicates.

Targeted sequencing and in silico analysis

An AmpliSeq Custom Panel was designed using Ion AmpliSeq Designer 2.2 (http://www.ampliseq.com/) against the exons of TP53, ESR1, PI3K, GATA3, MAP3K1 and CYP19A1 genes. For the preliminary analysis we filtered the design taking into consideration previously identified mutations (COSMIC, 141 amplicons, supplementary table 2). For the validation replicate we then included the entire set of exons (184 amplicons). Libraries were generated from 10 ng of DNA (tumour and normal) using the Ion AmpliSeq Library Kit v2.0 (Life Technologies) according to the manufacturer's instructions. Quantification of the libraries was performed using the Quant-iT dsDNA HS assay kit and a

Qubit2.0 fluorimeter (Life Technologies). Templates were prepared from a pool of equimolar amounts of each library using the Ion PGM Template OT2 200 Kit with OneTouch2 system (Life Technologies). Samples were sequenced on the Ion Torrent PGM™ sequencer using the Ion PGM™ 200 Sequencing Kit v2.0 on Ion 318™ chips. Data were analyzed using IonReporter and MuTect30 to compare metastatic samples with normal DNA (normal breast samples extracted from the same patient).

Fluorescence in situ hybridization (FISH) analysis

A FISH co-hybridization using a specific clone covering the altered locus and the specific alpha-satellite, as control probe for ploidy status, was performed on formalin-fixed paraffin embedded sections. Specifically, RP11-66L23 BAC clone for the CYP19A1 locus at 15q21.2 (red signal) and 15 Alpha Satellite probes (green signal) were used for the identification of CYP19A1 gene amplification. The BAC clone was selected using the University of California Santa Cruz Genome Browser Database (http://genome.ucsc.edu/) and was tested on normal human metaphase cells to verify the absence of cross-hybridization while the alpha-satellite probes were kindly provided by Dr. M. Rocchi, University of Bari, Italy. FISH experiments were performed as previously described, with minor modifications31. An average 30 representative nuclei scored per sample, scanning several areas to account for potential heterogeneity was counted to calculate the amplification ratio.

Single Cell RNA-FISH

The protocol for adherent mammalian cell lines was optimized for Stellaris FISH probes. Hybridization was performed overnight and no anti-fade was used for imaging. The sequence of the CAL Fluor Red 590 tagged probes targeting the CYP19A1 mRNA can be provided upon request. Samples were imaged using a Nikon Ti-E scanning laser confocal inverted microscope (A1) with 60x oil objective in tandem with Nikon NIS-Elements imaging software. Excitation was by 561.5 nm diode-pumped solid state. Detection was via 595-50 nm filter. Optical sections were captured at 0.300 μm intervals and a resolution of 256 by 256 pixels and zoom factor of 6.8, resulting in a voxel size of 0.0047 μm3 (0.1243 μm by 0.1243 μm by 0.3 μm). Four times averaging was used to reduce photon and camera noise. An automated spot count algorithm determined the number of mRNA 32. For the analysis we included 30 positive control cells to better define an mRNA spot.

Cell lines and hormone manipulation

Parental MCF7 breast cancer cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS). MCF7 cells and its derivatives were authenticated using STR profiling. Cells were routinely tested for Mycoplasma contamination. The chronically estrogen deprived MCF7-derived LTED breast cancer cell lines were maintained in phenol-red free DMEM containing 10% charcoal stripped fetal calf serum (SFCS). Both media were supplemented with 2 mM L-glutamine, 100 units/mL penicillin 0.1 mg/mL. Estradiol 10^{-8} M (E2758 Sigma) was added routinely to MCF7. Both cell lines were starved for 48h before further treatments. Subsequently, both cell lines were treated with 10^{-8} M Estradiol or Androstenedione (25nM final Concentration). LTED cells were also treated in the presence or absence of the Aromatase Inhibitor Letrozole (100nM

final concentration). After 24 hours, cells were lysed and RNA extracted with RNAeasy Micro Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. The quantity, quality and integrity of isolated mRNA were confirmed by absorption measurement and RNA gel electrophoresis. One (1) μg of RNA was retrotranscribed by using iSCRIPT (BioRad) containing random hexamers. Afterwards, quantitative real time PCR was carried out using SYBR select master mix (Life technologies, Paisley, UK), and ERα targets expression (TFF1, EGR3 and CA12) assessed. Values were quantified using the comparative threshold cycle method and target genes mRNA expression was normalized to GAPDH. Primers are available upon request. Results are expressed as means ± SEM. One-way Anova statistical analysis was performed using GraphPad Prism, and with GraphPad Software (GraphPad Software, Inc., Avenida de la Playa, La Jolla, USA). Two-sided p-values < 0.05 are considered statistically significant and are expressed as $*p < 0.05$. For CYP19A1 protein quantitation we used the Abcam ab71264 antibody. siRNA for CYP19A1 were obtained from ThermoFisher (siSilencer Select pre-validated s3875 and s3877). siRNA was transfected at 5nM final concentration two days prior to SRB analysis (day 0). SRB proliferation measurements were obtained after three further days of culture in the presence of increasing amount of letrozole. Experiments were conducted using five technical replicates and three independent biological replicates. CYP19A1 overexpressing cells were obtained by transfecting MCF7 cells with full length CYP19A1 (RC205890, OriGene Technologies) and selection using G418. SRB proliferation experiments were conducted as described above. For the exemestane challenge, MCF7 and LTED cells were plated in identical numbers and then treated with increased dose of exemestane (Tocris BioScience). SRB proliferation experiments were conducted as described above.

Patient-derived xenografts of ER+ breast cancer patients

Patient-derived xenografts were established from of ER+ metastatic breast cancer patients by injecting circulating and disseminated cancer cells isolated from the peripheral blood (CTC-) pleural effusion fluids (BPE-) or ascites (BA) into NOD.Cg-*PrkdcscidIl2rgtm1Wjl* mice, as described33,34. Analysis of DNA was performed from first or second passage xenograft tumors established in a mouse mammary fat pad. Animal care and all procedures were carried out according to German legal regulations and were previously approved by the governmental review board of the federal state of Baden-Württemberg, Germany (Regierungspräsidium Karlsruhe authorization number G240/11). This human material was obtained either from patients admitted to the University Clinic Mannheim Department of Gynecology or from patients recruited at the division of Gynecologic Oncology of the Heidelberg University Hospital. The study was approved by the ethics committee of the University of Heidelberg-Mannheim (case number 2011-380N-MA).

Survival Analysis

Kaplan-Meier plots were generated using PRISM (v5). We analyzed the time that separated the date of surgery from the date of first relapse (for all patients with univocal histology numbers). Data were analyzed using a Log-Rank Mantel-Cox test. Curves were also significantly different when analyzed using a Gehan-Breslow-Wilcoxon test.

Aromatase activity assay

Aromatase activity was evaluated using a 3 H- water release assay using 0.5 μmol/L of [1h-3 H]-androst-4-ene-3,17-dione as substrate35. The incubations were performed at 37 °C for 2 h under an air/CO2 (5%) atmosphere. The results obtained were expressed as fentomole-picomole/h and normalized to mg of protein (pmol/h/mg of protein).

CNV meta-analysis and Shallow-Sequencing

Meta-analysis of previously published data was conducted using cBioPortal (http://www.cbioportal.org/index.do). Amplification was scored as positive for GISTIC values of >2 (Amplified). SNPs array were interrogated using caSNP (http://cistrome.org/CaSNP/) using a CT threshold >3. Shallow-sequencing analysis were conducted using previously published data 12. Briefly, Input SAM files from resistant cell lines were used as ChIP tracks in MACS 1.4 against Input tracks generated in MCF7 cells. BED and WIG files were generated using default settings 36

ChIP-seq

ERα ChIP-seq data re-analyzed from previously published data 12. ERα bound regions were clustered using CHASE using K-means clustering (n=3) 37 (http://chase.cs.univie.ac.at/overview). RPKM plots were created for each specific cluster comparing tags from MCF7 and LTED cells. ChIP-seq data can be accessed under GEO profile GSE60517.

Statistical Methods

Mann-Whitney's, Student's T test (2-sided) and one way/two ways ANOVA with Bonferroni's posttest have been used as indicated in the figure legends. Assumptions on normal distribution and equal variance have been tested prior to statistical test's using ANOVA.

Supplementary Information

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976; 194:23–28. [PubMed: 959840]

2. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet. 2011; 378:771–784. [PubMed: 21802721]

3. Pagani O, et al. Adjuvant Exemestane with Ovarian Suppression in Premenopausal Breast Cancer. N Engl J Med. 2014; 371:107–118. [PubMed: 24881463]

4. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. Lancet. 2015; doi: 10.1016/S0140-6736(15)61074-1

5. Musgrove EA, Sutherland RL. Biological determinants of endocrine resistance in breast cancer. Nature Reviews Cancer. 2009; 9:631. [PubMed: 19701242]

6. Carroll JS, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell. 2005; 122:33–43. [PubMed: 16009131]

7. Magnani L, et al. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. Proc Natl Acad Sci USA. 2013; 110:E1490–9. [PubMed: 23576735]

8. Fuqua SA, Chamness GC, McGuire WL. Estrogen receptor mutations in breast cancer. J Cell Biochem. 1993; 51:135–139. [PubMed: 8440747]

9. Fribbens C, et al. Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. J Clin Oncol. 2016; JCO673061. doi: 10.1200/JCO.2016.67.3061

10. Robinson DR, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet. 2013; 45:1446–1451. [PubMed: 24185510]

11. Toy W, et al. ESR1 ligand-binding domain mutations in hormone- resistant breast cancer. Nature Genetics. 2013; 45:1439–1445. [PubMed: 24185512]

12. Nguyen VTM, et al. Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. Nature Communications. 2015; 6:10044.

13. Curtis C, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–352. [PubMed: 22522925]

14. Beroukhim R, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci USA. 2007; 104:20007–20012. [PubMed: 18077431]

15. Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6:pl1–pl1. [PubMed: 23550210]

16. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

17. Cao Q, et al. CaSNP: a database for interrogating copy number alterations of cancer genome from SNP array data. Nucleic Acids Res. 2011; 39:D968–D974. [PubMed: 20972221]

18. Koivisto P, et al. Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. Cancer Res. 1997; 57:314–319. [PubMed: 9000575]

19. Visakorpi T, et al. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet. 1995; 9:401–406. [PubMed: 7795646]

20. Koboldt DC, et al. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

21. Chen Z, Katzenellenbogen BS, Katzenellenbogen JA, Zhao H. Directed evolution of human estrogen receptor variants with significantly enhanced androgen specificity and affinity. Journal of Biological Chemistry. 2004; 279:33855–33864. [PubMed: 15159406]

22. Dunn CA, Clark W, Black EJ, Gillespie DAF. Estrogen receptor activation function 2 (AF-2) is essential for hormone-dependent transactivation and cell transformation induced by a v-Jun DNA binding domain-estrogen receptor chimera. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2003; 1628:147–155. [PubMed: 12932827]

23. Park J, Czapla L, Amaro RE. Molecular Simulations of Aromatase Reveal New Insights Into the Mechanism of Ligand Binding. J Chem Inf Model. 2013; 53:2047–2056. [PubMed: 23927370]

24. Ali S, Buluwela L, Coombes RC. Antiestrogens and Their Therapeutic Applications in Breast Cancer and Other Diseases. Annu Rev Med. 2011; 62:217–232. [PubMed: 21054173]

25. Cao Z, et al. Effects of resin or charcoal treatment on fetal bovine serum and bovine calf serum. Endocr Res. 2009; 34:101–108. [PubMed: 19878070]

26. Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. Genome Research. 2011; 21:940–951. [PubMed: 21460063]

27. Shaw LE, Sadler AJ, Pugazhendhi D, Darbre PD. Changes in oestrogen receptor-α and -β during progression to acquired resistance to tamoxifen and fulvestrant (Faslodex, ICI 182,780) in MCF7 human breast cancer cells. The Journal of Steroid Biochemistry and Molecular Biology. 2006; 99:19–32. [PubMed: 16533599]

28. Jeng MH. Estrogen Receptor Expression and Function in Long-Term Estrogen-Deprived Human Breast Cancer Cells. Endocrinology. 1998; 139:4164–4174.

29. Goldhirsch A, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer. Annals of Oncology. 2013; 24:2206–2223. [PubMed: 23917950]

30. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology. 2013; 31:213–219.

31. Pruneri G, et al. The transactivating isoforms of p63 are overexpressed in high-grade follicular lymphomas independent of the occurrence of p63 gene amplification. J Pathol. 2005; 206:337–345. [PubMed: 15887287]

32. Kempe H, Schwabe A, Crémazy F, Verschure PJ, Bruggeman FJ. The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. Mol Biol Cell. 2015; 26:797–804. [PubMed: 25518937]

33. Baccelli I, et al. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. Nature Biotechnology. 2013; 31:539–544.

34. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences. 2003; 100:3983–3988.

35. Lephart ED, Simpson ER. Assay of aromatase activity. Meth Enzymol. 1991; 206:477–483. [PubMed: 1784232]

36. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biology. 2008; 9:R137. [PubMed: 18798982]

37. Younesy H, et al. An Interactive Analysis and Exploration Tool for Epigenomic Data. Computer Graphics Forum. 2013; 32:91–100.
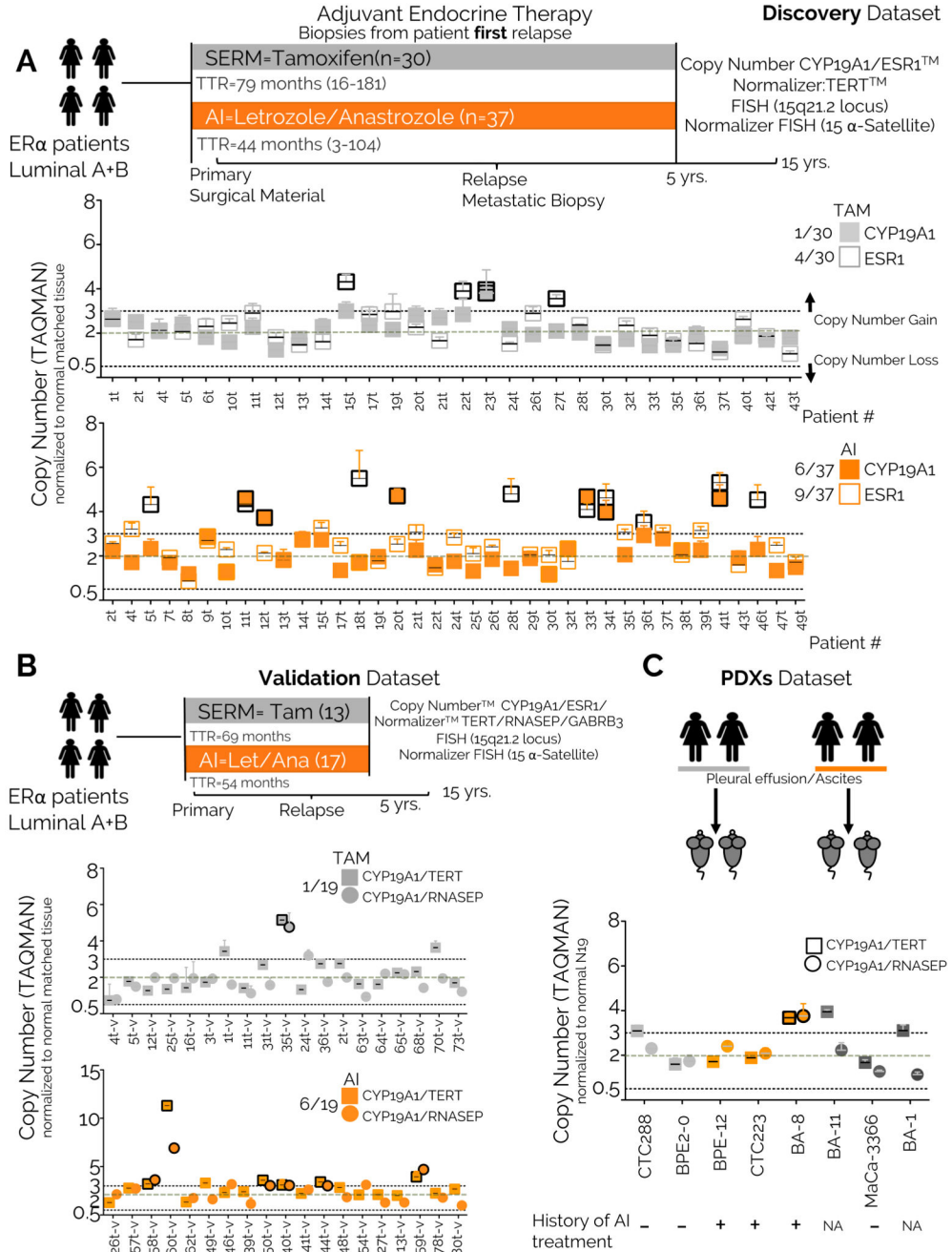
Figure 1. Clinical treatments shape cancer genetic evolution

A) Clinical discovery cohorts and sample design used in the study. CNA profiles for the *CYP19A1* and *ESR1* loci in the first relapse of patients treated with adjuvant Tamoxifen or AI mono-therapy B) Clinical discovery cohorts and sample design used in the study. CNA profiles for the *CYP19A1* and *ESR1* loci in the first relapse of patients treated with adjuvant Tamoxifen or AI mono-therapy. *ESR1* data can be found in Supplementary figure 4 C) PDXs cohort. CNA profiles for the *CYP19A1* and *ESR1* loci in PDXs from patient treated with Tamoxifen or AI. *ESR1* data can be found in Supplementary figure 4.
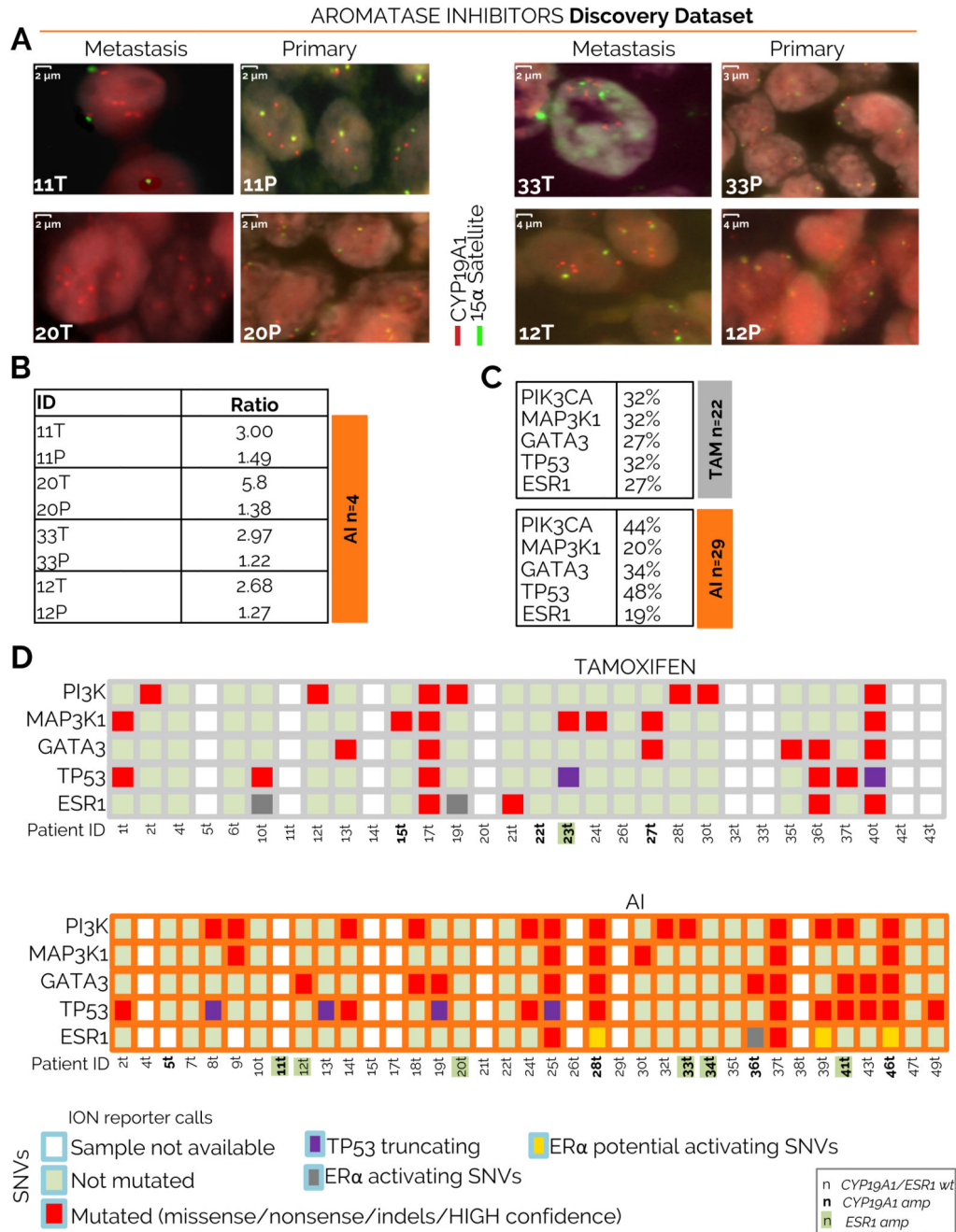
Figure 2. AI resistant metastases develop cluster amplification of the aromatase gene
A) Double-colour FISH analyses using 15 Alpha Satellite)/ *CYP19A1* probes identify cluster amplification of the *CYP19A1* locus B) Ratio of amplification obtained by computing CYP19A1/15α signals in 30 representative individual cancer cells from each validated tumor sample C) Breast Cancer mutations in Tamoxifen and AI treated metastatic samples D) Boxplots for individual patient/mutations in the two cohorts.
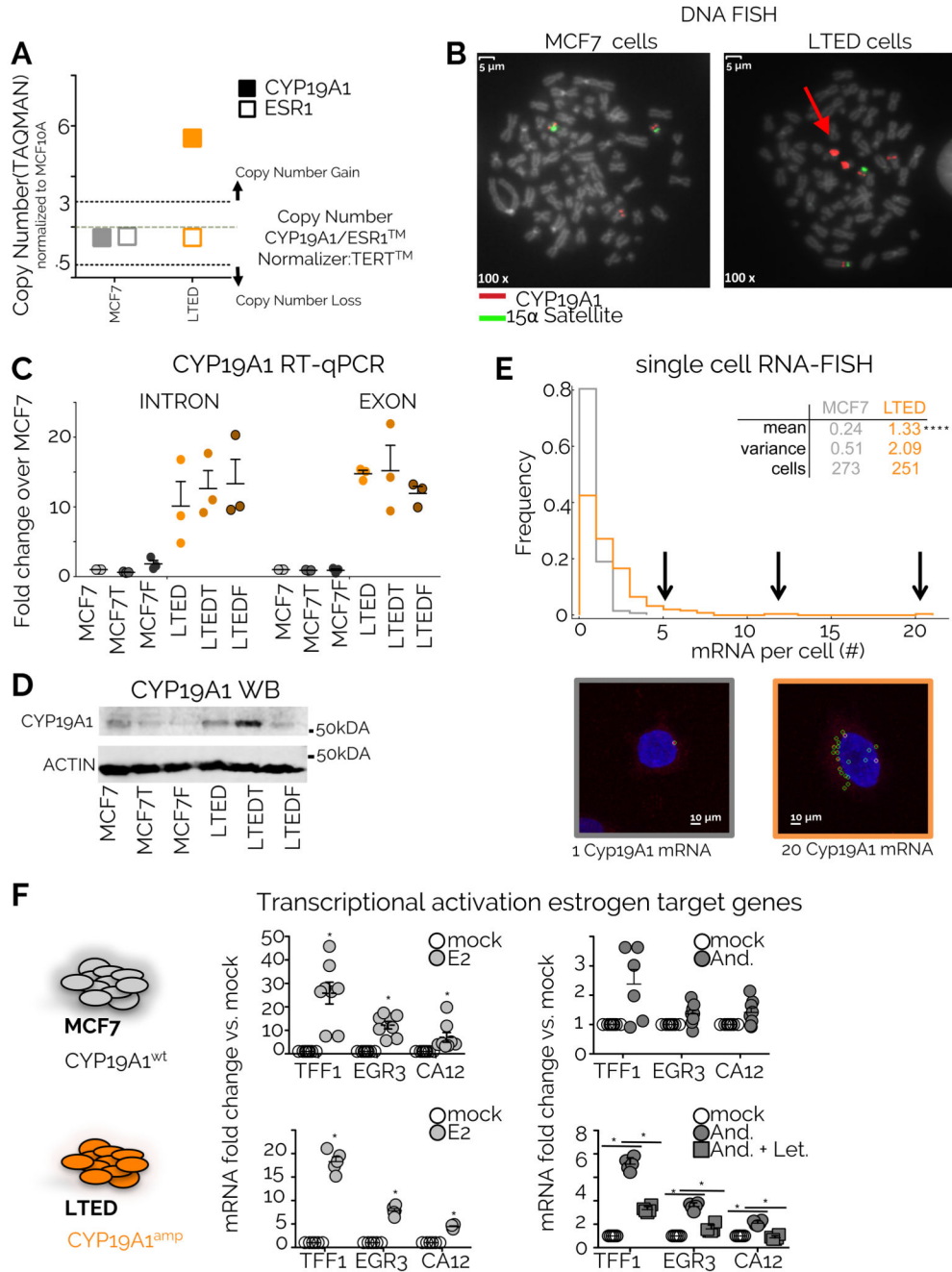
Figure 3. *CYP19A1* amplification leads to increased aromatase activity
A) CNA profiles for the *CYP19A1* and *ESR1* loci in treatment naïve (MCF7) and estrogen deprived (LTED) cells B) DNA-FISH using *CYP19A1*-centered probes identifies widespread cluster amplification in LTED cells C-D) LTED clones uniquely upregulate aromatase mRNA and protein levels. Dot-blots represent mean and S.E.M. from 3 independent experiments. Western Blot have been cropped near the specific band. Full blot can be found in Supplementary Figure 11 E) Single-cell RNA-FISH highlight heterogeneity in aromatase expression. Asterisks denote a significant difference after Mann-Whitney test

****=P<10^{-20} F) CYP19A1^{amp} cells transcriptionally activate estrogen-target genes in response to aromatizable androgens. AI treatment blocks transcriptional activation. Dot-blots represent mean and S.E.M. from 5 independent experiments. Asterisks denote a significant difference after two-way Anova or one-way ANOVA (bottom right panel) *=P<0.05.
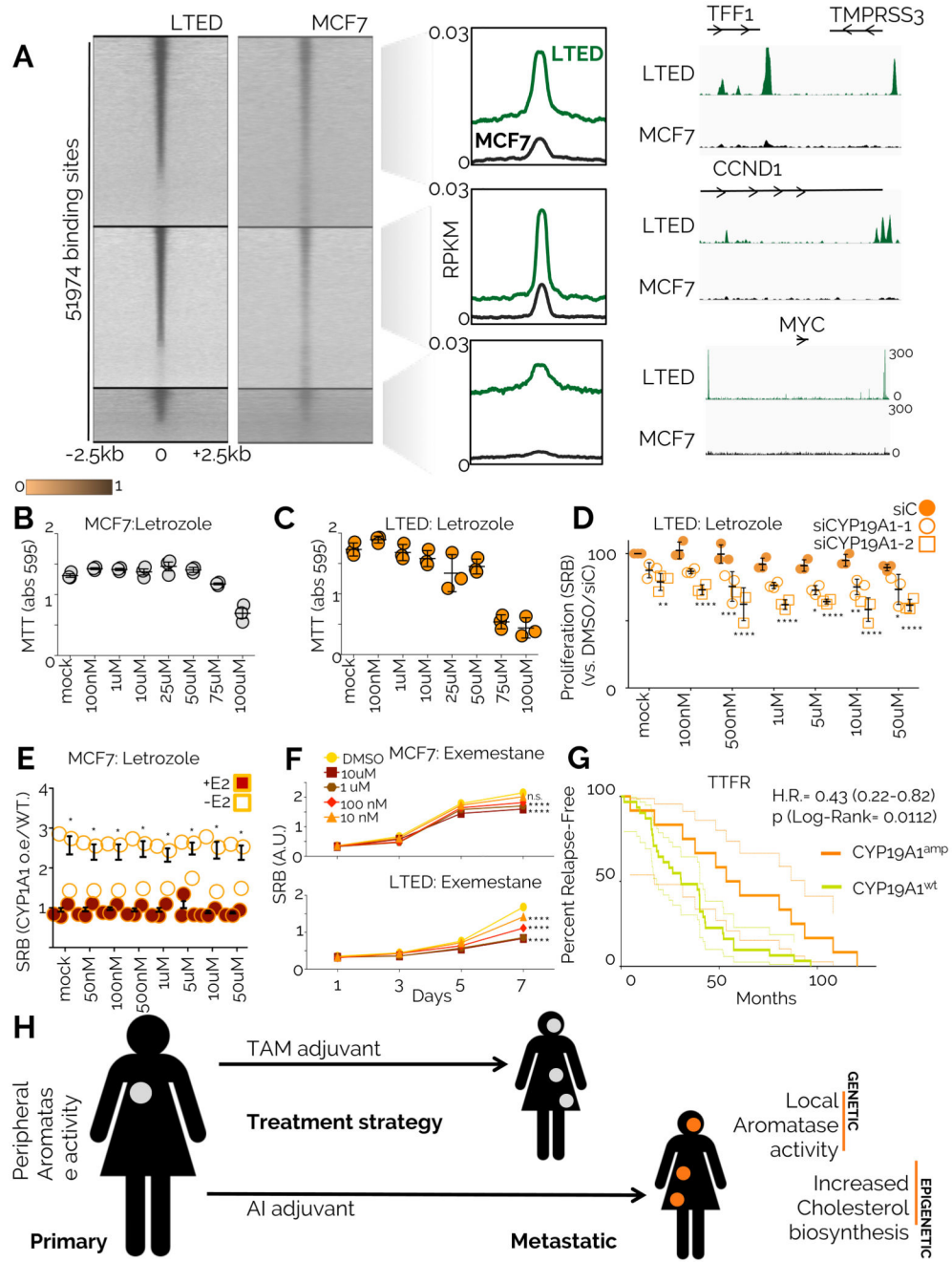
Figure 4. *CYP19A1^amp* cells endogenously activate ERα and develop tolerance to AI
A) ChIP-seq heatmaps for ERα in treatment naïve (MCF7) and estrogen deprived (LTED) cells. Binding sites have been assigned to three clusters. The average profile of each cluster is reported in the central panels. Examples of ERα enrichment near important estrogen target genes are shown in the insets (right panels). B) MCF7 treatment with AI in the presence of estradiol. Dot-blots represent mean and S.E.M. from 3 independent experiments. C) LTED treatment with AI in the absence of estradiol. Dot-blots represent mean and S.D. from 3 independent experiments. Asterisks denote a significant difference after Student T-

test *=P<0.05) D) CYP19A1 depleted LTED cells have increased sensitivity to AI. Dot-blots represent mean and S.D. from 3 independent experiments. Asterisks denote a significant difference after two-way ANOVA and Bonferroni post-test *,**,***,****= P<0.05, 0.01, 0.001 and 0.0001 E) CYP19A1 over-expressing cells have a growth advantage compared to WT in the absence of estradiol. Relative increase in growth rate is shown by plotting the ratio between the growth of CYP19A1 over-expressing cells to CYP19A1 WT cells under letrozole challenge. Dot-blots represent mean and S.E.M. from 3 independent experiments. Asterisks denote a significant difference after two-way ANOVA and Bonferroni post-test *=P<0.05 F) *CYP19A1^{amp}* LTED respond to low levels of irreversible steroidal AI. Dots represent mean and 95% C.I of 4 independent replicates. Asterisks denote a significant difference after two-way ANOVA and Bonferroni post-test ****= P<0.0001 G) Kaplan-Meier curve showing time to first relapse (TTF) for AI treated patients stratified retrospectively for CYP19A1 amplification. Dotted lines represent the 95% confidence intervals H) Working hypothesis for therapy-specific breast cancer progression. Genetic and epigenetic changes collaborate to increase tumor fitness by creating an estrogen-independent niche at metastatic sites treated with AI therapy.