Activating ESR1 mutations in hormone-resistant metastatic breast cancer

Dan R Robinson1,2,12, Yi-Mi Wu1,2,12, Pankaj Vats1,2, Fengyun Su1,2, Robert J Lonigro1,3, Xuhong Cao1,4, Shanker Kalyana-Sundaram1,2, Rui Wang1,2, Yu Ning1,2, Lynda Hodges1, Amy Gursky1,2, Javed Siddiqui1,2, Scott A Tomlins1,2, Sameek Roychowdhury5, Kenneth J Pienta6, Scott Y Kim7, J Scott Roberts8, James M Rae3,9, Catherine H Van Poznak9, Daniel F Hayes9, Rashmi Chugh9, Lakshmi P Kunju1,2, Moshe Talpaz9, Anne F Schott9 & Arul M Chinnaiyan1–4,10,11

Breast cancer is the most prevalent cancer in women, and over two-thirds of cases express estrogen receptor-α (ER-α, encoded by ESR1). Through a prospective clinical sequencing program for advanced cancers, we enrolled 11 patients with ER-positive metastatic breast cancer. Whole-exome and transcriptome analysis showed that six cases harbored mutations of ESR1 affecting its ligand-binding domain (LBD), all of whom had been treated with anti-estrogens and estrogen deprivation therapies. A survey of The Cancer Genome Atlas (TCGA) identified four endometrial cancers with similar mutations of ESR1. The five new LBD-localized ESR1 mutations identified here (encoding p.Leu336Gln, p.Tyr537Ser, p.Tyr537Cys, p.Tyr537Asn and p.Asp538Gly) were shown to result in constitutive activity and continued responsiveness to anti-estrogen therapies in vitro. Taken together, these studies suggest that activating mutations in ESR1 are a key mechanism in acquired endocrine resistance in breast cancer therapy.

Advances in high-throughput sequencing technologies are beginning to establish a molecular taxonomy for a spectrum of human diseases and has facilitated a move toward ‘precision medicine’ (refs 1,2). With regard to oncology, defining the mutational landscape of a patient’s tumor will lead to more precise treatment and management of individuals with cancer. Comprehensive clinical sequencing programs for cancer patients have been initiated at a variety of medical centers, including our own3,4. In addition to the potential for identifying ‘actionable’ therapeutic targets in cancer patients, these clinical sequencing efforts may also shed light on acquired resistance mechanisms developed against targeted therapies5–7.

ER is the primary therapeutic target in breast cancer and is expressed in 70% of cases8. Drugs directly antagonizing ER, such as tamoxifen and fulvestrant, are a mainstay of breast cancer treatment; however, approximately 30% of ER-positive breast cancers exhibit de novo resistance, whereas 40% acquire resistance to these therapies9. In addition to anti-estrogen therapies, patients with ER-positive breast cancer are also treated with aromatase inhibitors such as letrozole and exemestane10. Aromatase inhibitors block the peripheral conversion of androgens into estrogen and, in post-menopausal women, lead to over a 98% decrease in circulating levels of estrogen. As with anti-estrogens, treatment with aromatase inhibitors results in the development of resistance, but this is presumably due to different mechanisms, as patients with breast cancer who develop resistance to aromatase inhibitors often still respond to anti-estrogen therapies11. The molecular mechanisms of endocrine resistance in ER-positive breast cancer continues to be an active area of research12.

Our institutional review board (IRB)-approved clinical sequencing program, called MI-ONCOSEQ (the Michigan Oncology Sequencing Program), enrolls patients with advanced cancer across all histologies3. Since April 2011, we have enrolled over 200 patients in this program, which involves obtaining a current tumor biopsy with matched normal samples (blood and/or buccal swab). Samples are then subjected to integrative sequencing, which includes whole-exome sequencing of the tumor and matched normal sample, transcriptome sequencing and, as needed, low-pass whole-genome sequencing1. This combination of DNA and RNA sequencing technologies allows one to be relatively comprehensive with regard to the mutational landscape of coding genes, including analysis of point mutations, indels, amplifications, deletions, gene fusions or translocations, and outlier gene expression profiles. These results are generated within a 5- to 7-week time frame and are presented at an institutional ‘precision medicine tumor board’ to deliberate upon potentially actionable findings.

1Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, Michigan, USA. 2Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan, USA. 3Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan, USA. 4Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, Michigan, USA. 5Department of Internal Medicine, Ohio State University, Columbus, Ohio, USA. 6Brady Urological Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. 7Center for Bioethics and Social Science in Medicine, University of Michigan, Ann Arbor, Michigan, USA. 8Department of Health Behavior & Health Education, University of Michigan, Ann Arbor, Michigan, USA. 9Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA. 10Department of Urology, University of Michigan Medical School, Ann Arbor, Michigan, USA. 11Center for Computational Medicine and Biology, University of Michigan, Ann Arbor, Michigan, USA. 12These authors contributed equally to this work. Correspondence should be addressed to A.M.C. (arul@umich.edu).

Received 22 July; accepted 10 October; published online 3 November 2013; doi:10.1038/ng.2823
As part of the MI-ONCOSEQ program, we enrolled and sequenced 11 patients with metastatic ER-positive breast cancer (Table 1 and Supplementary Table 1). A diverse array of aberrations were identified in individual patients, some of which are potentially actionable, including mutations in PIK3CA (n = 4), BRCA1 aberrations (n = 2), FGFR2 aberrations (n = 2)\(^\text{13}\), NOTCH2 frameshift deletion (n = 1), cyclin and associated cyclin-dependent kinase aberrations (n = 3) and MDM2 amplification and overexpression (n = 1). Aberrations were also frequently found in the tumor suppressor TP53 (n = 6), the DNA mismatch repair gene MSH2 (n = 1) and in epigenetic regulators (n = 2), including ARID2, ARID1A and SMARCA4, among others. The complete spectra of somatic mutations with associated alterations in expression levels and copy number in the index patients are given in Supplementary Figure 1 and Supplementary Tables 2 and 3. Two of the index patients, MO_1031 and MO_1051, exhibited a high level of mutations consistent with ‘signature B’ identified in a whole-genome study of mutational processes in breast cancer\(^\text{14}\). There were 39 gene fusions identified in the 6 index patients, with 11 encoding in-frame fusion proteins (Supplementary Fig. 2 and Supplementary Tables 4 and 5), including an activating FGFR2-AFF3 fusion\(^\text{13}\).

The most notable observation in the mutational landscapes of these treated patients with ER-positive breast cancer was the finding of nonsynonymous somatic mutations in ESR1 affecting the LBD (n = 6). The six index patients MO_1031, MO_1051, MO_1069, MO_1129, MO_1167 and MO_1185 had mutations encoding p.Leu536Gln, p.Asp538Gly, p.Tyr537Ser and p.Asp538Gly and p.Tyr537Ser alterations in the LBD, respectively. The respective mutation in each case was detected by whole-exome sequencing of the tumor relative to the matched normal sample and was corroborated by whole-transcriptome sequencing, as ESR1 was expressed at moderate to high levels (Supplementary Table 2). The clinical histories of the index patients are depicted in timelines in Figure 1. For three of the patients (MO_1051, MO_1069 and MO_1129), we had access to primary diagnostic material and showed that the ESR1 mutations were not present at an earlier stage, indicating that they were acquired after endocrine therapy (Fig. 1 and Supplementary Table 2). Interestingly, all of the index patients were treated with anti-estrogens (tamoxifen and/or fulvestrant) and aromatase inhibitors (letrozole, anastrozole and/or exemestane). Two of the patients also had an oophorectomy. Comparison of the mutations present in each primary versus post-treatment pair showed a substantial number of shared mutations in both samples of the pair, including activating mutations in PIK3CA in two of the cases. Thus, it is clear that the index cases presented with recurrent disease of the original primary tumor surviving in an estrogen-deprived state and having acquired ESR1 mutations. Of note, neither ESR1 amplifications nor gene fusions were observed in these cases.

The five new LBD alterations of ESR1 identified in this study are depicted in Figure 2. Each occurred in the vicinity of the synthetic alterations of ESR1 that are inverted in response to tamoxifen and involve p.Met543Ala and p.Leu544Ala alterations (Inv-mut-AA2)\(^\text{15}\) and served as a positive control for our subsequent in vitro studies. We next investigated the occurrence of ESR1 mutations in a range of breast cancer types. Here we took advantage of data from the TCGA Project, which has generated whole-exome sequences for 27 tumor types in at least 4,000 individual samples. As expected, LBD-disrupting mutations of ESR1 were not detected in the 390 ER-positive breast cancers sequenced by TCGA, as these were primary resection samples before hormonal treatment\(^\text{16}\), nor did we detect ESR1 mutations in a cohort of 80 triple-negative breast carcinoma transcriptomes (D.R.R., Y.-M.W., X.C., S.K.-S., A.M.C. et al., unpublished data).

As the LBD-disrupting mutations of ESR1 we identified were somatic and were acquired after treatment, we next assessed whether the encoded proteins were dependent on estrogen for activation. We cloned into expression vectors each of the five ESR1 mutants identified in this study (encoding p.Leu536Gln, p.Tyr537Ser, p.Asp538Gly, p.Tyr537Cys and p.Tyr537Asn alterations) and subsequently cotransfected these constructs into HEK293T cells with an estrogen response element (ERE)-luciferase reporter system. We then exposed steroid hormone-deprived cells to β-estradiol for 24 h and assessed ERE reporter levels. Surprisingly, unlike wild-type ESR1, which had little ERE reporter activity in the absence of ligand, all five of the ESR1 mutants had strong constitutive activation of the ERE reporter that was not markedly enhanced with β-estradiol (Fig. 3). This finding suggested that each of the mutations developed in the context of evolution during...
an estrogen-deprived state. Consistent with this idea, a whole-genome sequencing study of 46 patients with ER-positive breast cancer enrolled in 2 aromatase inhibitor trials did not identify any of these ESR1 mutations in the pretreatment samples analyzed\textsuperscript{17}.

Next, we assessed whether anti-estrogen therapies affected the functional activity of the LBD mutants. As effects on inhibition can be determined, we performed a dose response study with expression plasmid and selected a dose of 50 ng for the following experiments\textsuperscript{18} (Supplementary Fig. 3). As expected, wild-type ESR1 was inhibited in a dose-dependent fashion by the anti-estrogens 4-hydroxytamoxifen, fulvestrant and endoxifen (Fig. 4 and Supplementary Figs. 4–6). In addition, the mutant corresponding to the synthetic ESR1 mutation (Inv-mut-AA2) was activated in a dose-dependent fashion by these anti-estrogens (Fig. 4), which has been reported previously\textsuperscript{15}. Interestingly, ESR1 with each of the five LBD alterations identified in this study was inhibited by tamoxifen and fulvestrant in a dose-dependent fashion and did not exhibit the inverted response to anti-estrogens that the synthetic Inv-mut-AA2 mutant did. One could speculate that the corresponding mutations did not arise under selective pressure of anti-estrogen treatment but rather in the context of an estrogen deprivation setting, such as treatment with aromatase inhibitors and/or oophorectomy. The IC\textsubscript{50} (half-maximal

Figure 1 Clinical timelines for the six index ER-positive metastatic breast cancer patients harboring ESR1 mutations. Shown are patients’ histories of clinical treatment from first diagnosis until enrollment in the MI-ONCOSEQ study. Each bar represents the timeframe of treatment. FFPE, formalin fixed, paraffin embedded; CAF, cyclophosphamide, doxorubicin, fluorouracil chemotherapy.

Supplementary Figs. 4–6). In addition, the mutant corresponding to the synthetic ESR1 mutation (Inv-mut-AA2) was activated in a dose-dependent fashion by these anti-estrogens (Fig. 4), which has been reported previously\textsuperscript{15}. Interestingly, ESR1 with each of the five LBD alterations identified in this study was inhibited by tamoxifen and fulvestrant in a dose-dependent fashion and did not exhibit the inverted response to anti-estrogens that the synthetic Inv-mut-AA2 mutant did. One could speculate that the corresponding mutations did not arise under selective pressure of anti-estrogen treatment but rather in the context of an estrogen deprivation setting, such as treatment with aromatase inhibitors and/or oophorectomy. The IC\textsubscript{50} (half-maximal

Figure 1 Clinical timelines for the six index ER-positive metastatic breast cancer patients harboring ESR1 mutations. Shown are patients’ histories of clinical treatment from first diagnosis until enrollment in the MI-ONCOSEQ study. Each bar represents the timeframe of treatment. FFPE, formalin fixed, paraffin embedded; CAF, cyclophosphamide, doxorubicin, fluorouracil chemotherapy.

Supplementary Figs. 4–6). In addition, the mutant corresponding to the synthetic ESR1 mutation (Inv-mut-AA2) was activated in a dose-dependent fashion by these anti-estrogens (Fig. 4), which has been reported previously\textsuperscript{15}. Interestingly, ESR1 with each of the five LBD alterations identified in this study was inhibited by tamoxifen and fulvestrant in a dose-dependent fashion and did not exhibit the inverted response to anti-estrogens that the synthetic Inv-mut-AA2 mutant did. One could speculate that the corresponding mutations did not arise under selective pressure of anti-estrogen treatment but rather in the context of an estrogen deprivation setting, such as treatment with aromatase inhibitors and/or oophorectomy. The IC\textsubscript{50} (half-maximal
inhibitory concentration) values for both 4-hydroxytamoxifen and fulvestrant were two- to fourfold higher for all mutants compared to wild-type ESR1. Fulvestrant exhibited greater maximal inhibition than 4-hydroxytamoxifen for all the mutants tested (Supplementary Figs. 4 and 5).

The ESR1 alterations identified in this study cluster near the beginning of helix 12 (Fig. 2). Structural studies have demonstrated a key role for the position of helix 12 in the response of the ER to agonists and antagonists,12, and Tyr537 has been postulated to form a capping motif contributing to the activity of the receptor.20 Specifically, the p.Tyr537Ser mutant has been reported to have higher affinity for estrogen than wild-type ESR1 and interacts with the SRC1 coactivator in the absence of ligand.21 Several studies using experimental mutagenesis have implicated the same three residues identified here as critical determinants of the transcriptional activity of the receptor.21,23,24

As estrogen therapy has been shown to have a positive effect in treating aromatase inhibitor–resistant advanced breast cancers, we tested the effect of low- to high-dose estrogen on the activity of the mutants in the transient luciferase reporter assays (Supplementary Fig. 7)25,26. The results did not suggest that the effectiveness of this therapy is mediated through direct control of the transcriptional activity of these mutants, if encoded by the responding patients.

Although the primary intent of our broad-based clinical sequencing program is to identify actionable and/or driver mutations in advanced cancers, this study demonstrates how such prospective, real-time sequencing efforts can also shed light on resistance mechanisms that develop against targeted therapies. A number of resistance mechanisms have been suggested to function in the evasion of endocrine treatment, including activation of the mTOR and phosphoinositide 3-kinase (PI3K) pathways, among others.9,27. Although the total number of ER-positive breast cancers we have sequenced is modest, we have done so in a comprehensive fashion in terms of delineating mutational landscapes and incorporating both DNA and RNA sequencing. This analysis identified de novo driver mutations and/or potentially acquired mutations in breast cancer such as mutations resulting in PI3K activation, PAK1 amplification and FGFR fusion and amplification, which have been described previously.13,20,28 Among potential new mechanisms described, we identified profound focal amplification of MDM2 (which encodes a negative regulator of p53 that is targetable) and copy gains of GNRHR (encoding gonadotropin-releasing hormone receptor), which may be related to past endocrine therapy.

As the LBD-disrupting mutations of ESR1 identified in this study result in constitutive activity, the encoded mutant proteins can function in the absence of ligand and maintain ER signaling. In 1997, an ESR1 mutation affecting the LBD, encoding a p.Tyr537Asn alteration, was detected in a single individual with stage IV metastatic breast cancer who had been treated with diethylstilbestrol, but, since then, this mutation has been considered to be very rare.29,30 With the advent of widespread aromatase inhibitor therapy, we suggest that alteration of the ESR1 LBD is likely a common mechanism of resistance that develops in low-estrogen states. Interestingly, LBD-disrupting mutations of ESR1 were detected somatically in 4 of 373 cases of endometrial cancer.31 We speculate that the four TCGA endometrial tumors that harbor LBD-affecting mutations likely came from patients with concurrent breast cancer, as tamoxifen treatment is known to be associated with higher incidence of this tumor type and such patients also often receive estrogen deprivation treatment.32

Our study suggests that it is unlikely that these LBD alterations develop in the context of anti-estrogen treatment, as the mutated ESR1 variants continue to be responsive to direct ER antagonists such as tamoxifen and fulvestrant. This finding is consistent with clinical reports showing that patients that develop resistance to aromatase inhibitors still respond to anti-estrogen treatment.31 Although this prospective clinical sequencing study was not designed to characterize a specific cancer type or treatment resistance mechanism,
Figure 4  Acquired ESR1 mutations result in maintained sensitivity to anti-estrogen therapies. HEK293T cells were cotransfected with an ERE–firefly luciferase reporter plasmid, a plasmid constitutively expressing Renilla luciferase and various ESR1 constructs as indicated. (a,b) Steroid hormone–deprived cells were either untreated or treated with increasing doses of the anti-estrogen drugs tamoxifen (a) and fulvestrant (b) in the presence of 5 nM β-estradiol (E2) for 24 h. Percentage change in ESR1-driven transcriptional activity was calculated using E2-treated cells as the control for each tested construct. Data shown are the means from triplicate experiments. Error bars, s.d. *P < 0.001. Red, wild type; blue, clinically identified mutants; orange, synthetic ligand-inversion mutant; green, vector control.

future studies comprising larger cohorts of breast cancer patients with disease that recurs after varied endocrine treatments will more precisely delineate the incidence of this acquired resistance mechanism. The focused nature of these mutations and their role in aromatase inhibitor resistance suggest the possibility of monitoring patients undergoing treatment using circulating tumor DNA methods33,34. In this manner, treatment could be shifted to head off evolving tumor resistance.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence data have been deposited in the database of Genotypes and Phenotypes (dbGaP), which is hosted by the National Center for Biotechnology Information (NCBI), under accessions phs000602.v1.p1 and phs000673.v1.p1.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
The authors thank D. Miller, T. Barrette and D. Gibbs for hardware and database management, K. Giles for assistance with manuscript preparation, physicians M. Wicha, L. Pierce, D. Smith, K. Levin and F. Feng for referring patients, and C. Betts and J. Athanikar for assistance with tumor boards. We also thank the larger contributors to this project, including L. Higgins, E. Caoili and R. Dunnick. This project is supported in part by the Prostate Cancer Foundation, the Alfred A. Taubman Institute, the American Cancer Society, the Howard Hughes Medical Institute and a Doris Duke Charitable Foundation Clinical Scientist Award.

AUTHOR CONTRIBUTIONS

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

A mutational processes molding the genomes of 21 breast α endometrial cancer in tamoxifen-treated breast cancer patients: Nature America, Inc. All rights reserved.
ONLINE METHODS

Clinical study and specimen collection. Sequencing of clinical samples was performed under IRB-approved studies at the University of Michigan. Patients were enrolled and consented for integrative tumor sequencing in MI-ONCOSSEQ (Michigan Oncology Sequencing Protocol, HUM00046018). Medically qualified patients 18 years or older with advanced or refractory cancer were eligible for the study. Informed consent detailed the risks of integrative sequencing and includes up-front genetic counseling. Informed consent was obtained from all subjects included in this study. Biopsies were arranged for safely accessible tumor sites. Needle biopsies were snap frozen in OCT (Optimal Cutting Temperature) compound, and a longitudinal section was cut. Frozen sections stained with hematoxylin and eosin were reviewed by pathologists to identify cores with the highest tumor content. Remaining portions of each needle biopsy core were retained for nucleic acid extraction.

Extraction of DNA and RNA. Genomic DNA from frozen needle biopsies and blood was isolated using the Qiagen DNeasy Blood and Tissue kit, according to the manufacturer's instructions. Total RNA was extracted from frozen needle biopsies using the Qiazol reagent with disruption using a 5-mm bead on a Tissuelyser II (Qiagen) and was purified using a miRNasey kit (Qiagen) with DNAse I digestion, according to the manufacturer's instructions. RNA integrity was verified on an Agilent 2100 Bioanalyzer using RNA Nano reagents (Agilent Technologies).

Preparation of next-generation sequencing libraries. Transcriptome libraries were prepared using 1–2 μg of total RNA. Polyadenylated RNA was isolated using Sera-Mag oligo(dT) beads (ThermoScientific) and fragmented with the Ambion Fragmentation Reagents kit. cDNA synthesis, end repair, A-base addition and ligation of the Illumina indexed adapters were performed according to Illumina's TruSeq RNA protocol. Libraries were selected for DNA fragments of 250–300 bp in size on a 3% Nusieve 3:1 agarose gel (Lonza), recovered using QIAEX II gel-extraction reagents (Qiagen) and PCR amplified using Phusion DNA polymerase (New England BioLabs). Amplified libraries were purified using AMPure XP beads (Beckman Coulter). Library quality was measured on an Agilent 2100 Bioanalyzer by product size and concentration. Paired-end libraries were sequenced with the Illumina HiSeq 2000 platform (2 × 100-nucleotide read length). Reads that passed the chastity filter of Illumina BaseCall software were used for subsequent analysis.

Exome libraries of matched pairs of tumor and normal genomic DNA were generated using the Illumina TruSeq DNA Sample Prep kit, following the manufacturer's instructions. In brief, 1–3 μg of each genomic DNA sample was sheared using a Covaris S2 to a peak target size of 250 bp. Fragmented DNA was concentrated using AMPure XP beads, and end repair, A-base addition and ligation of Illumina indexed adapters were performed. Adapter-ligated libraries were electrophoresed on 3% Nusieve agarose gels, and fragments of 300–350 bp were recovered using QIAEX II gel-extraction reagents. Recovered DNA was amplified using Illumina index primers for eight cycles and purified using AMPure XP beads, and DNA concentration was determined using a Nanodrop spectrophotometer. Libraries (1 μg) were hybridized to the Agilent SureSelect Human All Exon v4 chip at 65 °C for 60 h, following the manufacturer's protocol (Agilent Technologies). Targeted exon fragments were captured on Dynal M-280 streptavidin beads (Invitrogen) and enriched by amplification with the Illumina index primers for nine additional PCR cycles. PCR products were purified with AMPure XP beads and analyzed for quality and quantity using an Agilent 2100 Bioanalyzer and DNA 1000 reagents.

We used the publicly available software FastQC to assess sequencing quality. For each lane, we examined per-base quality scores across the length of the reads. Lanes were deemed passing if the per-base quality score box plot indicated that >85% of the reads had >Q20 for bases 1–100. In addition to raw sequence quality, we also assessed alignment quality using the Picard package. This allows monitoring of duplication rates and chimeric reads that may result from ligation artifacts, crucial statistics for interpreting the results of copy number and structural variant analysis.

Gene fusion detection. Paired-end transcriptome sequencing reads were aligned to the human reference genome (GRCh37/hg19) using an RNA sequencing (RNA-seq) spliced read mapper TopHat2 (ref. 35) (TopHat 2.0.4) with the ‘–fusion-search’ option turned on to detect potential gene fusion transcripts. In the initial process, TopHat2 internally deploys an ultrafast short-read alignment tool, Bowtie (Version 0.12.8), to map the transcriptome data. Potential false-positive fusion candidates were filtered out using the ‘TopHat-Post-Fusion’ module. Further, fusion candidates were manually examined for annotation and ligation artifacts. Junction reads supporting the fusion candidates were realigned using the BLAT alignment tool to confirm fusion breakpoints. Full-length sequence of each fusion gene was constructed on the basis of supporting junction reads and evaluated for potential ORFs using an ORF Finder. For gene fusions with robust ORFs, the amino acid sequences of the fused proteins were explored using the Simple Modular Architecture Research Tool (SMART) to examine the gain or loss of known functional domains in the fusion proteins.

Gene expression. BAM ‘accepted_hits.bam’ files, which were generated by the TopHat mapping module, were used to quantify the expression data through Cufflinks (Version 2.0.2), an isoform assembly and RNA-seq quantification package. The structural features of 56,369 transcripts from the Ensembl resource (Ensembl 66) were used as an annotation reference to quantify the expression of individual transcripts and isoforms. The ‘Max Bundle Length’ parameter was set to ‘10000000’, and ‘multi-read-correct’ was flagged on to perform an initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome.

Mutation analysis. Whole-exome sequencing was performed on an Illumina HiSeq 2000 or HiSeq 2500 instrument in paired-end mode, and primary base call files were converted into FASTQ sequence files using the bc12fastq converter tool bc12fastq-1.8.4 in the CASAVA 1.8 pipeline. FASTQ sequence files were then processed through an in-house pipeline constructed for whole-exome sequence analyses of paired cancer and normal genomes. Sequencing reads were aligned to reference genome build hg19 (GRCh37) using Novoalign multithreaded (Version 2.08.02, Novocraft) and converted into BAM files using SAMtools (Version 0.1.18)(37). Sorting and indexing of BAM files used Novosort threaded (Version 1.00.01), and duplicate reads were removed using Picard (Version 1.74). Mutation analysis was performed using VarScan2 algorithms (Version 2.3.2) (38) with the pileup files created by SAMTools mpileup for tumor and matched normal samples, simultaneously performing pairwise comparisons of base call and normalized sequence depth at each position. For SVN detection, filtering parameters including coverage, variant read support, variant frequency, P value, base quality, the presence of homopolymers and strandedness were applied. For indel analysis, Pindel (Version 0.2.4) was used on tumor and matched normal samples, and indels common to both samples were classified as germline, whereas indels present in tumor but not in normal samples were classified as somatic. Finally, a list of candidate indels as well as of somatic and/or germline mutations was generated by excluding synonymous SNVs. ANNOVAR (39) was used to functionally annotate the detected genetic variants, and positions are based on Ensembl 66 transcript sequences.

Tumor content for each tumor exome library was estimated from the sequence data by fitting a binomial mixture model with two components to the set of most likely SVN candidates from two-copy genomic regions. The set of candidates used for estimation consisted of coding variants that (i) were supported by at least 3 variant fragments in the tumor sample, (ii) were not supported by variant fragments in the matched benign sample, with at least 16 fragments of coverage, (iii) were not present in dbSNP, (iv) were within a targeted exon or within 100 bp of a targeted exon, (v) were not in homopolymer runs of 4 or more bases and (vi) exhibited no evidence of amplification or deletion. To filter out regions of possible amplification or deletion, we used exon coverage ratios to infer copy number changes, as described below. Resulting SVN candidates were not used for the estimation of tumor content if the segmented log ratio exceeded 0.2 in absolute value. Candidates on the Y chromosome were also eliminated because they were unlikely to exist in two-copy genomic regions. Using this set of candidates, we fit a binomial mixture model with two components using the R package flexmix, version 2.3-8. One component consisted of SVN candidates with very low variant fractions, presumably resulting from recurrent sequencing errors and other artifacts. The other component, consisting of the set of likely true SNVs, was informative of tumor content in the tumor sample. Specifically,
under the assumption that most or all of the observed SNV candidates in this component are heterozygous SNVs, we expect the estimated binomial proportion of this component to represent one-half of the proportion of tumor cells in the sample. Thus, the estimated binomial proportion obtained from the mixture model was doubled to obtain an estimate of tumor content.

Copy number aberrations were quantified and reported for each gene as the segmented, normalized, log$_2$-transformed exon coverage ratio between each tumor sample and its matched normal sample. To account for observed associations between coverage ratios and variation in GC content across the genome, lowess normalization was used to correct per-exon coverage ratios before segmentation analysis. Specifically, mean GC percentage was computed for each targeted region, and a lowess curve was fit to the scatterplot of log$_2$ coverage ratios versus mean GC content across the targeted exome using the lowess function in R (version 2.13.1) with smoothing parameter $f = 0.05$.

Partially redundant sequencing of areas of the genome affords the ability for cross-validation of findings. We cross-validated exome-based point mutation calls by manually examining the genomic and transcriptomic reads covering the mutation using the UCSC Genome Browser. Likewise, gene fusion calls from the transcriptome data can be further supported by structural variant detection in the genomic sequence data, as well as by copy number information derived from genome and exome sequencing.

Chemicals and reagents. β-estradiol, (Z)-4-hydroxytamoxifen, (E/Z)-endoxifen hydrochloride hydrate and fulvestrant were purchased from Sigma-Aldrich.

Plasmids and cloning. cDNA for wild-type ESR1 was PCR amplified from a breast cell line MCF7 (ATCC) with the introduction of a sequence encoding an N-terminal Flag tag. cDNAs encoding the relevant mutations of ESR1 were generated by site-directed mutagenesis (QuikChange, Agilent Technologies), and full-length constructs were fully sequenced. All ESR1 variants were placed in the lentiviral vector pCDH (System Biosciences) for eukaryotic expression.

ERE-luciferase reporter assays. For cell transfection experiments, HEK293T cells (ATCC) were plated at a density of 1–2 × 10$^5$ cells per well (24-well plates) in phenol red–free DMEM containing 10% FBS and antibiotics. Once cells attached, the medium was replaced with DMEM containing 10% charcoal/dextran-treated FBS (HyClone), and cells were cultured overnight. The next day, cells were transiently cotransfected with ESR1 expression plasmid (50 ng/well) and luciferase reporter constructs (25 ng/well; SABiosciences) using FuGene 6 reagent (Promega). The ER-responsive luciferase plasmid encoding the firefly luciferase reporter gene is driven by a minimal CMV promoter and tandem repeats of the estrogen transcriptional response element (ERE). A second plasmid constitutively expressing Renilla luciferase served as an internal control for normalizing transfection efficiencies (Signal ERE Reporter, SABiosciences). After transfection for 18 h, cells were serum starved for a few hours before treatment with β-estradiol or anti-estrogen drugs. Cells were harvested 18 h after treatment, and luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega). IC$_{50}$ values were computed using the GraphPad Prism application to fit a four-parameter dose response curve.