HHS Public Access

Author manuscript

Clin Cancer Res. Author manuscript; available in PMC 2019 October 31.

Published in final edited form as:

Clin Cancer Res. 2018 October 01; 24(19): 4887–4899. doi:10.1158/1078-0432.CCR-17-3702.

Comprehensive profiling of DNA repair defects in breast cancer identifies a novel class of endocrine therapy resistance drivers

M Anurag[1,2], **N Punturi**[1,2], **J Hoog**[3], **MN Bainbridge**[4,5], **MJ Ellis**[6,2,*], **S Haricharan**[6,2,*]

[1]Department of Medicine, Baylor College of Medicine, Houston, Texas.

[2]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, Texas.

[3]Siteman Cancer Center Breast Cancer Program, Washington University School of Medicine, St. Louis, Missouri.

[4]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.

[5]Rady Children's Institute for Genomic Medicine, San Diego, California.

[6]Department of Medicine, Baylor College of Medicine, Houston, Texas.

Abstract

Purpose—This study was undertaken to conduct a comprehensive investigation of the role of DNA damage repair (DDR) defects in poor outcome ER+ disease.

Experimental design—Expression and mutational status of DDR genes in ER+ breast tumors were correlated with proliferative response in neoadjuvant aromatase inhibitor therapy trials (discovery data set), with outcomes in METABRIC, TCGA and Loi data sets (validation data sets), and in patient derived xenografts. A causal relationship between candidate DDR genes and endocrine treatment response, and the underlying mechanism, was then tested in ER+ breast cancer cell lines.

Results—Correlations between loss of expression of three genes: *CETN2* ($p<0.001$) and *ERCC1* ($p=0.01$) from the nucleotide excision repair (NER) and *NEIL2* ($p=0.04$) from the base excision repair (BER) pathways were associated with endocrine treatment resistance in discovery data sets, and subsequently validated in independent patient cohorts. Complementary mutation analysis supported associations between mutations in NER and BER pathways and reduced endocrine treatment response. A causal role for *CETN2*, *NEIL2* and *ERCC1* loss in intrinsic endocrine resistance was experimentally validated in ER+ breast cancer cell lines, and in ER+ patient-derived xenograft models. Loss of *CETN2*, *NEIL2* or *ERCC1* induced endocrine treatment response by dysregulating G1/S transition, and therefore, increased sensitivity to CDK4/6 inhibitors. A combined DDR signature score was developed that predicted poor outcome in multiple patient cohorts.

Conclusion—This report identifies DDR defects as a new class of endocrine treatment resistance drivers and indicates new avenues for predicting efficacy of CDK4/6 inhibition in the adjuvant treatment setting.

*Corresponding authors: Svasti.Haricharan@bcm.edu, Matthew.Ellis@bcm.edu.

Keywords

Endocrine therapy resistance; DNA damage repair; Estrogen-receptor positive; breast cancer; Genomics; mutations; transcriptomics; mRNA expression; Ki67; AIinhibitors; prognosis; CETN2; ERCC1; NEIL2; PRKDC

INTRODUCTION

Breast cancer is the most frequent form of cancer affecting women, and estrogen-receptor positive (ER+) tumors account for 60–70% of all reported cases. For patients with early stage ER+ disease, endocrine therapy: tamoxifen or an aromatase inhibitor (AI) are preferred first-line therapies. Despite these treatments, at least 1 in 4 patients develop fatal endocrine therapy resistance (1, 2). Although some markers predictive of endocrine treatment (ET) response are available, for example *ERBB2* mutation/amplification (3, 4), gene expression profiles (5) and on-neoadjuvant endocrine treatment Ki67 analysis (6, 7), resistance largely remains an unpredictable and poorly understood event. Efforts to study underlying mechanisms of ET resistance have focused on activation of peptide growth factors (e.g. *EGFR, ERBB2*) and on activating mutations or translocation in *ESR1*(8, 9). However, these mechanisms mostly explain adaptive or acquired resistance to endocrine treatment in the advanced setting, thereby circumscribing their use as predictive markers in primary tumors. The success of cyclin dependent kinase (CDK) 4/6 inhibition for metastatic breast cancer (10, 11) indicates a major role for the target cell cycle dependent kinases in restoring growth control in ET resistant tumors, but these agents lack accurate preductive markers. Consequently adjuvant treatment will lead to over-treatment in some patients and under-treatment in others.

It is known that loss of the MutL components of the mismatch repair (MMR) complex causes poor initial response to ET (intrinsic ET resistance) but in the MutL deficient setting tumors remain sensitive to CDK4/6 inhibition (12). Effects of other DNA damage response/repair (DDR) defects on ET resistance are understudied but might have similar relationships. In other cancer types, disruptions in DDR pathways associate with tumor formation, responsiveness to chemotherapy, and loss of replicative checkpoints in many cancer types(13). Additionally, ER-induced signaling and proliferation downregulates many DDR pathways in the normal mammary gland(14), potentially signifying a relationship between defects in DDR and hormonal response in ER+ breast cancer cells. Several genomic studies on breast cancer have identified signatures of DNA repair defects generated by classifying types of mutations, but the impact of these studies have been diluted by uncertainty regarding the molecular origin and clinical relevance of these signatures (15). Therefore, there is strong rationale for conducting a comprehensive, molecular analysis of the role for DDR defects in regulating and predicting ET response.

DDR is constituted of 8 canonical pathways: mismatch repair (MMR) – which can be further broken down into MutL and MutS complementation groups -- nucleotide excision repair (NER), base excision repair (BER), non-homologous end joining (NHEJ), homologous recombination (HR), Fanconi Anemia (FA), trans-lesion synthesis (TLS) and direct repair (DR)(16). The first five of these pathways fall into one of two larger categories:

single strand break repair (SSBR) consisting of MMR, BER and NER, and double strand break repair (DSBR) comprised of NHEJ and HR(17) (Figure 1A). Here, we describe a comprehensive analysis of these canonical DDR pathways in ER+ breast tumors from patients and draw causal associations between BER and NER defects and ET resistance that parallel earlier observations on the role of MMR defects.

METHODS

DDR genes set compilation

Gene set for eight canonical DDR pathways comprising of Direct Repair (DR), Mismatch repair (MMR), Nucleotide excision repair (NER), Base excision repair (BER), homologous recombination (HR), non-homologous end joining (NHEJ) and Fanconi Anemia (FA) along with checkpoint genes was built as a union of MSigDB (18, 19) KEGG (c2:curated) pathway specific genes and updated table of DDR pathway genes listed at http://sciencepark.mdanderson.org/labs/wood/dna_repair_genes.html. Genes shared across different DDR pathways or checkpoints were not included in the analysis. Additionally, Translesion DNA synthesis (TLS) pathway genes were left out from current analysis, because of its ambiguous role in SSBR and DSBR mechanism. Cytoscape (20) was used to visualize the DDR pathway network.

Patient data

Data sets—Z1031/POL data set (referred to as NeoAI) was used with permission from the Alliance consortium (21). The data was obtained after written informed consent from the patients and the studies were conducted in accordance with recognized ethical guidelines approved by IRB (21). TCGA (downloaded: 06/17) and MSKCC-IMPACT (downloaded: 02/18) mutation data were obtained from cBio portal. TCGA (downloaded: 06/17) and METABRIC (downloaded: 06/17) copy number data were obtained from cBio portal. TCGA analyses were restricted to ER+ patient tumors (except for Figure S3 where basal-like and HER2-enriched tumors were analyzed based on published PAM50 categorization(22)). For MSKCC-IMPACT analyses, a list of ER− breast cancer sample IDs was derived from previously published literature (23) and subtracted from the list available on cBio Portal. While there is estimated to be some contamination from HER2+ or ER− breast cancer patients, this subtracted list contains a majority of ER+ breast tumors. TCGA and METABRIC gene expression data and associated survival outcomes were downloaded from Oncomine. Standard cut-offs of mean-1.5x standard deviation were used to identify “Low” subsets of each candidate gene in each data set when multiple candidate genes were combinatorially analyzed. For individual analyses in METABRIC, “Low” subsets were identified using median cut-offs, and in Loi, using mean-1.5x standard deviation.

Enrichment analysis—Mutations from NeoAI (21, 24) and were used as the discovery set along with DNA-sequenced ER+ tumors from TCGA. For NeoAI data set, mutations with Normal VAF<10/NA and Tumors VAF >10/NA were classified as somatics. Mutations were scored by SIFT (25, 26) to assess the effect on the protein structure and function. In accordance with SIFT standards, missense mutations with scores <0.05 were considered to be damaging. For the enrichment analysis in TCGA, mutations were categorized into three

variant types – Missense, frameshift and nonsense. Frameshift and nonsense mutations were cumulatively referred to as FS/NS mutations. Enrichment analysis in TCGA used the Z-score test of two population proportions to compare the proportion of missense to frameshift/nonsense mutations in each DDR pathway to the proportion of missense to frameshift/nonsense mutations in a control set of unrelated genes (*MYH7*, *SYNE1*, *NEB*) in tumors from patients who remained alive. This approach was used to ensure the presence of appropriate sample size for the analysis employed and to control for genome-wide levels of mutations in each tumor.

Survival analysis—For the univariate and multivariate analysis, we analyzed 887 tumors from Luminal (A/B) patients who received endocrine treatment. mRNA (microarray) expression and survival information along with other clinical metadata were extracted from Oncomine (27–29). Only samples with survival metadata were included in the analysis. For this cohort, tumors with expression level of candidate genes lower than median values were labeled as “low” while rest were labelled as “high”. All survival data was analyzed using Kaplan-Meier curves and log-rank tests. Proportional hazards were determined using Cox regression.

Statistical Analysis

Missing data were imputed with “NA” from mutation, expression and survival data analysis. Samples classifying for more than one category were either removed (e.g. samples with mutations in both SSBR and DSBR genes) or treated as separate set (e.g. cumulative analysis of tumors with dysregulation of new versus published candidates) for statistical comparisons. Pearson’s correlation was performed on the log transformed normalized data for every DDR gene using automated script in R. To control errors in multiple test corrections, false discovery rates were calculated using the Benjamini-Hochberg (30) method in R. Two-tailed Wilcoxon rank sum tests were used for two-sample tests of association between classes. Pathway over-representation analysis was performed using thirteen candidate genes as the input genes against all DDR genes (n=104) as the background in WebGestalt (31) using KEGG pathway database.

An expression signature score known as CENMP (CETN2,ERCC1, NEIL2, MLH1, PMS2) score was devised using mean of standardized expression for the mentioned fives genes. The CENMP score was calculated for the three independent data set (NeoAI, Metabric and Loi) using the microarray expression levels. Survival of patients with highest 20% quantile of CENMP score was compared against that of patients with lowest 20% quantile of the same.

Cell lines, siRNA transfection and growth assays

ZR75.1 and MCF7 parental cells were from ATCC, 2015 and 2017 respectively, and tested for mycoplasma contamination upon arrival using the Lonza Mycoalert Plus Kit (CAT# LT07–710) as per the manufacturer’s instructions, and annually since. Both lines were maintained in RPMI 1640 1X w/ L-glutamine and 1% Penicillin/Streptomycin (Sigma-Aldrich, CAT#P4333–100mL). Cell lines used for experiments were <20 passages. Transient transfections with esiRNA (Sigma-Aldrich) towards human NEIL2 (CAT#EHU158461), CETN2 (CAT#EHU137031), ERCC1 (CAT#EHU156971), RAD23B (CAT#EHU145881)

and POLM (CAT#SASI_Hs02_003344553), or scrambled control used at 50nmoles/L each were delivered using Polypluys jetPRIME® transfection reagent (CAT#114–07) as per the manufacturer's instructions. Cells were plated for experiments 48hrs after transfection. Stable selection with puromycin after infection with lentivirus harboring RNAi oligos (ABM) against human NEIL2 (CAT#i014980a), CETN2 (CAT#i004479a), and ERCC1 (CAT#i007023a) or scrambled control (CAT#i000238c) was conducted using manufacturer's protocol. Growth assays were repeated independently in triplicate as previously reported (32) using Alamar blue to detect cell viability. Final readings were between 4–6 days after initial drug treatment and fold change plotted for analysis.

Drug-treatment

Fulvestrant(Fisher Scientific CAT#506242), 4-OHT(Sigma-Aldrich CAT#H7904), Palbociclib (ThermoFisher CAT#508548) and Abemaciclib (ThermoFisher CAT#NC0577560), dissolved in DMSO, and β-estradiol (Sigma-Aldrich CAT#E2758–1G), dissolved in water, at 10mmoles/L were stored long term at −80C with working stocks at −20C. Cells were treated 24 hours after plating with fresh drug added every 48hrs until the end of the experiment. Estradiol addition experiments were conducted in phenol red-free DMEM with 10% charcoal stripped serum.

qPCRs, Western blot, FACS and IFs

RNA from cell lines was extracted using the Qiagen RNeasy Mini Kit (CAT#74106) and converted to cDNA using Bio-Rad Reverse Transcriptase iScript (CAT#TX1708891BAY), both following manufacturers' instructions. cDNA was quantified by qPCR using Bio-Rad SsoAdvanced Universal SYBR Green supermix (CAT#17525272) at the manufacturers specifications. For Immunoflourescence, 48hrs after esiRNA transfection cells were plated on Poly-D-Lysine coated coverslips (CAT#NC0746078), treated for 24hrs, then harvested. Cells on coverslips were then washed with 1x PBS, fixed in 4% PFA, and then co-stained with Ki-67/MKI67 Antibody (Novus-Biologicals CAT#NB110–89717SS). Western blotting was conducted as described previously (32) using antibodies against CETN2 (ABclonal CAT#A5397), NEIL2 (Abcam CAT#ab221556), ERCC1 (Abcam CAT#ab129267), and β-Actin (Sigma-Aldrich CAT#A5316–100ul). FACS analysis for PCNA (Abcam CAT#ab29) was conducted after treating MCF7 siScr or siCEN cells with 100nM Fulvestrant for 72 hours following a 20-hour treatment with 250nM nocodazole (Sigma-Aldrich CAT#M1404). Cells were then probed for PCNA and run on a BD Accuri C6 flow cytometer using a standard protocol.

RESULTS

NER and BER downregulation associates with endocrine treatment resistance

The status of 104 DDR genes belonging to the six major DDR pathways: NER, BER, MMR, NHEJ, FA and HR was assessed in primary ER+ breast tumors. Genes that were shared between multiple pathways were excluded from the analysis to facilitate better understanding of the discrete contributions of individual pathways to ET resistance (Figure 1A). Additionally, direct repair (DR) and trans-lesion synthesis (TLS) were excluded because they fall outside canonical SSBR-DSBR categorization, and appeared to be rarely

dysregulated in ER+ tumors in initial observations. The discovery set, referred to as NeoAI, was composed of data from two neoadjuvant aromatase inhibitor trials (Z1031(21, 33) and POL(24)). These clinical trials were designed to assess intrinsic ET response by accruing serial biopsies from patients at diagnosis (baseline, BL), after 2–4 weeks of endocrine treatment (on-treatment) and in the surgical specimen (Figure 1B). These biopsies were annotated by whole genome/exome sequencing, RNA-seq and gene expression microarray, providing both mutational and transcriptomic information. The biopsies, both baseline and on-treatment were also evaluated for Ki67, a marker of proliferation, by both immunohistochemistry (IHC) and expression(34) (Figure S1A). An on-treatment Ki67 value >10% is a clinically relevant marker of intrinsic ET resistance associated with elevated risk of relapse in the first 5 years of follow up (35). The discovery strategy for this study was therefore, to correlate DDR gene expression at baseline with on-treatment Ki67 levels (both by IHC and by expression), and combine results with that from a complementary enrichment analysis for deleterious mutations, to identify a set of DDR genes which when dysregulated may predict intrinsic ET resistance (Figure 1B).

Transcriptomic analysis (schema outlined in Figure S1B) identified 13 DDR genes whose reduced expression (at baseline) correlated with increased on-treatment Ki67 levels (Figure 2A). Three genes (*ESR1*, *GATA3* and *RUNX1*)(36–39) were used as positive controls because of their known associations with ET responsiveness (Figure 2A). Twelve of the 13 DDR genes identified belonged to SSBR pathways, corresponding to ~20% of all unique SSBR genes compared to only 2% of all unique DSBR genes (Figure 2B) supporting a strong role for SSBR pathway downregulation in ET resistance. FA was the only one of three DSBR pathways examined to demonstrate significant association with ET resistance (Figure 2C). Pathway enrichment analysis of the candidate gene list relative to all other DDR genes studied revealed significant over-representation of “Platinum drug resistance” (p=0.04, comprised of MMR and NER genes) and “mismatch repair” (p=0.04) terms (Figure 2D). These results framed the proposition that underexpression of genes serving NER, and to a lesser extent, BER genes can reduce response to ET.

Low-expression of *CETN2*, *ERCC1* and *NEIL2* are poor prognostic factors in ER+ breast cancer patients treated with endocrine therapy

To understand the effect of downregulation of the 13 candidate genes on long-term patient outcomes, univariate associations between low-expression of these 13 genes and patient survival was tested in an independent data set (METABRIC). To increase the specificity of these correlations only the subset of patients with luminal tumors who were treated with ET was included in these analyses. Univariate Cox Proportional hazard analyses based on disease-specific survival identified five genes as significantly associating with poor survival: *CETN2*, *ERCC1*, *MLH1*, *NEIL2* and *PMS2* (Figure S2). Subsequent multivariate analyses, including tumor size, grade and node status, supported an independent role for these five candidates in predicting disease-specific survival (Figure 3A). An association between reduced expression of MutL genes, *MLH1* and *PMS2,* and poor survival has already been described in this data set (12). Kaplan-Meier survival analysis demonstrated that low expression of *CETN2*, *NEIL2*, and *ERCC1* individually also correlated with worse disease-specific survival compared to all other ER+ breast cancer patients (Figure 3B–D). However,

no association between low RNA levels and poor survival for *CETN2, NEIL2,* and *ERCC1* was observed in either HER2-enriched (Figure S3A–C) or basal-like (Figure S3D–F) breast cancer patients, suggesting that the association between defects in these genes and survival is subtype-specific. A correlation between low expression of *CETN2*, *NEIL2* and *ERCC1* and poor recurrence-free survival was also observed in the Loi data set (40), serving as independent validation (Figure S4). Further, a composite signature based on low expression of any one of these genes (the CEN signature) associated with a significantly increased risk ratio of 5.1 in Loi (p=0.02, Figure 3E).

Damaging mutations in nucleotide excision repair, base excision repair and non-homologous end joining genes are enriched in ER+ patient tumors

To further understand the involvement of DDR genes in ER+ breast cancer, incidence of damaging vs. non-damaging missense mutations (as predicted by SIFT (25, 26)) was analyzed in pre-treatment biopsies from NeoAI (Table S1). In this analysis, genes from SSBR pathways showed significant enrichment for damaging mutations when compared to all other genes in the genome whereas DSBR genes did not (Figure 4A). Enrichment for damaging and tolerant mutations over genome-wide frequency was then assessed for each individual DDR pathway in NeoAI (Figure S5). Damaging mutations were enriched in genes of NER, BER, NHEJ and HR pathways (FDR<0.05), but tolerant mutations were not, potentially indicating a selection for deleterious mutations in these pathways during tumor evolution. Damaging mutations were also enriched in genes of FA pathway but non-damaging mutations showed even higher enrichment, suggesting that any role for FA gene mutation in ER+ breast cancer is likely complex (Figure S5). Enrichment for deleterious (frameshift/nonsense; FS/NS) over missense (MS) mutations in ER+ patient tumors was validated in TCGA for genes of BER and NHEJ, but not HR and FA, pathways, although due to limited follow-up time in this data set similar validation could not be obtained for NER (Figure 4B). To facilitate rigorous statistical analysis, p-values were generated by comparing the proportion of somatic FS/NS:MS mutations in each DDR pathway in patients who were alive, to a control set of genes that have not been implicated as cancer drivers (*SYNE1*, *MYH7*, *NEB*). Together, these enrichment analyses promote the postulate that NER, BER and NHEJ gene mutations may be ER+ breast cancer drivers.

To address associations with clinical outcomes more directly, Cox Regression analysis was conducted for tumors with mutations in each DDR pathway in two data sets: TCGA and MSKCC-IMPACT (Figure 4C). TCGA has whole exome sequence data from >800 ER+ breast tumors, while MSKCC-IMPACT has targeted sequencing of a selected panel of genes (including a subset of DDR genes) in >300 ER+ primary breast tumors. Amongst the SSBR pathways, mutations in NER (*ERCC2–5*) and BER genes each associated with significantly higher hazard ratio in MSKCC-IMPACT and TCGA databases, respectively (Figure 4C), validating observations made in the gene expression (Figure 2) and enrichment analyses described above (Figure 4A&B; Figure S5). Association of NER gene mutations in TCGA and BER gene mutations in MSKCC-IMPACT could not be made because of median follow-up being <6 months in either case.

Amongst the DSBR pathways, tumors with mutations in NHEJ genes associated with a significantly higher hazard ratio when compared to wildtype tumors in TCGA (Figure 4C). No NHEJ gene was included in the targeted panel sequenced in MSKCC-IMPACT precluding validation in this data set. To date, NHEJ has not been associated with ET resistance. Only five genes from this pathway had truncating mutations in either NeoAI or TCGA: *PRKDC, XRCC5, DNTT, NHEJ1* and *POLM*. Patients whose tumors harbored either MS or FS/NS mutations in any of these genes associated with worse overall survival in TCGA, as did patients whose tumors had copy number loss of these loci (Figure S6A). When individual associations with survival were analyzed, only mutation and/or copy number loss of *PRKDC* associated independently with poor survival (HR=2.8, p=0.009, Figure S6B). Additionally, tumors with either *PRKDC* copy number loss or mutations also had significantly lower gene expression of *PRKDC* ($p<0.001$, Figure S6C), suggesting that this is unlikely to be a chance association. Although mutation data on *PRKDC* was not available in METABRIC, the association of *PRKDC* copy number loss with poor survival was validated in METABRIC (Figure S6D). Additionally, the association of *PRKDC* mutations with poor prognosis was recently observed in an independent set of ER+ patient tumors (41).

Amongst the other DSBR pathways, primary tumors with mutations in HR genes did not associate with higher hazard ratios than wildtype tumors in either TCGA or MSKCC-IMPACT where eight HR genes (*RAD54L*, *RAD52*, *RAD51D*, *RAD51*, *NBN*, *BRCA1* and *BLM*) were included in the targeted panel (Figure 4C). Mutations in FA genes had mixed associations, with patients whose tumors had mutations in FA genes associating with worse survival in TCGA but not in MSKCC-IMPACT (*FANCA*, *FANCC*, *PALB2*, *BRIP1*) (Figure 4C).

Consideration of all three discovery parameters analyzed, i.e. gene expression downregulation, gene mutation, and association with patient outcomes (increased hazard ratios for overall survival in TCGA, METABRIC or MSKCC-IMPACT) provides strongest support for the involvement of SSBR pathway dysregulation in poor clinical outcomes of ER + breast cancer patients (Figure 4D). More specifically, all three discovery parameters support an understudied role for NER and BER dysregulation in ET resistance (Figure 4D) that warrants functional investigation. Evidence for involvement of DSBR pathway dysregulation in ER+ breast cancer outcomes is less consistent across the three different screening parameters (Figure 4D) and requires further investigation in larger patient cohorts, and in experimental model systems.

Two confounding factors affect interpretation of mutational analyses conducted here. Firstly, it is possible that dysregulation of replication factors commonly associated with DSBR disruption affects proliferative response to ET. However, no decisive association between replication gene expression (*RPA1–4*) and on-treatment proliferation marker Ki67 (IHC and mRNA, Figure S7) was observed in NeoAI, suggesting that replicative disruption is unlikely to be a major confounding factor for this analysis. Secondly, previous reports have suggested an association between high mutation load or genome instability and poor patient outcomes in breast cancer (42), which may influence interpretation of associations between mutations in DDR genes and clinical outcome. No significant increase in genome instability or

mutation load was observed in tumors with mutations in DSBR genes (Figure S8B&D). However, as expected, a significant increase in mutation load, but not genome instability, was observed in SSBR mutated tumors (Figure S8A&C). Therefore, it is not possible to rule out high mutation load as a potential confounding factor in the mutational analysis presented above, without functional validation of causality. Dysregulation of genes from the NER (*CETN2*&*ERCC1*) and BER (*NEIL2*) pathways were, therefore, functionally investigated since candidates from these pathways were most consistently correlated with ET resistance and poor patient outcomes (Figure 4D).

Experimental validation of *CETN2*, *ERCC1* and *NEIL2* as endocrine therapy resistance genes

To test whether dysregulation of candidate genes from NER and BER pathways can directly cause ET resistance, pooled siRNA against each of the three candidate genes identified in the gene expression screen, i.e. *CETN2*, *ERCC1* and *NEIL2*, as well as a scrambled control, was transiently transfected into two ET-sensitive, ER+ breast cancer cell lines, MCF7 (Figure 5) and ZR-75 (Figure S9). Knockdown of the genes was validated at RNA level in both cell lines (Figure S9A&C), and downregulation of each protein was confirmed by Western blots of MCF7 cell lysates (Figure 5A). Cells transfected with siRNA against scrambled control (siScr), *CETN2*, *NEIL2* or *ERCC1* were then exposed to all three classes of ET: estrogen deprivation in media containing charcoal stripped serum (to mimic AI), and tamoxifen or fulvestrant treatment in media containing full serum. Estrogen deprived siCETN2, siNEIL2, and siERCC1 MCF7 (Figure S9B) and ZR-75 (Figure S9D) cells showed attenulated growth response to estradiol stimulation when compared to siScr control cells, indicating decreased influence of estrogen signaling on proliferaton. Consistent with this notion, siCETN2, siNEIL2 and siERCC1 MCF7 (Figure 5B&C) and ZR-75 (Figure S9E&F) cells demonstrated a significant lack of growth inhibition in response to either fulvestrant or tamoxifen treatment.

As negative controls, two DDR genes that did not correlate with ET resistance, *RAD23B* and *POLM*, were also knocked down using siRNA in MCF7 cells (Figure S10A). Downregulation of these genes did not alter growth response to either fulvestrant (Figure S10B) or tamoxifen (Figure S10C). Additionally, independent lentiviral RNAi oligos against *CETN2*, *NEIL2* and *ERCC1* were used to stably select MCF7 cells with *CETN2*, *ERCC1* and *NEIL2* knockdown (Figure S9A&S10D demonstrate knockdown at RNA and protein level respectively). Fulvestrant-independent (Figure 10E) and tamoxifen-independent (Figure S10F) growth phenotypes were faithfully replicated in these stable cells, providing orthogonal confirmation that growth effects caused by knockdown of the three candidate genes, *CETN2*, *NEIL2* and *ERCC1* are specific and causal.

Next, dysregulation of these candidates at either the mutational or RNA level was assessed across 7 ER+ PDXs (BCaPE(43)). One of the 7 lines had strong downregulation of *NEIL2* expression, and two other lines exhibited downregulation of *ERCC1*. Additionally, one line had downregulation of *PMS2* (MutL-). One of the lines with low *ERCC1* RNA also harbored a missense mutation in *MLH1* suggesting a compound phenotype. All four of these lines, (designated CENMP- because of a disruption of any CETN2, ERCC1, NEIL2 or MutL

component) exhibited significantly higher tumor viability after treatment with the AI, anastrozole, when compared with the other three tumors designated CENMP+ in this PDX cohort (p=0.02, Figure 5D).

To test whether the loss of proliferative inhibition to ET observed in human tumors (Figure 2) was reproducible in experimental systems, Ki67 levels were also assessed before and after fulvestrant treatment in siCETN2, siNEIL2 and siERCC1 MCF7 cells relative to siScr control. Inhibition of gene expression from any of these three candidates resulted in a profound lack of Ki67 inhibition in response to fulvestrant treatment, unlike control cells (Figure S11A), reproducing observations in clinical trial samples. Earlier studies indicated that MutL-defective ER+ breast cancer cells exhibit altered proliferative response to ET due to dysregulation of G1/S cell cycle transition (12). To test whether the candidate DDR genes identified in this screen also participate in the regulation of G1/S transition, their gene expression pattern was analyzed across the cell cycle after a double thymidine block (www.dnarepairgenes.com (44)). All three candidates, as well as NHEJ genes, had maximal gene expression specifically in G1 or around the G1/S transition point, similarly to *MLH1* (Figure 5E). On the other hand, FA gene expression was maximal in late S phase and HR gene expression in late G2 (Figure 5E). These observations are consistent with published data (45, 46) and indicate a common role for the candidate endocrine resistance DDR genes identified here in G1/S transition. As in the case of MutL-defective tumors, CEN/PRKDC-ER+ patient tumors from TCGA also had significantly increased RNA levels of *CDK4*, the principal G1 cyclin-dependent kinase (Figure S11B) and protein levels of PCNA, a marker of successful S phase transition, relative to CEN/PRKDC+ tumors (Figure S11C). Increased PCNA positivity was also confirmed in MCF7 cells with stable knockdown of *CETN2*, *NEIL2*, *ERCC1* and *MLH1* after fulvestrant treatment relative to control cells (Figure S11D).

To test ER regulation as an alternative mechanism uniting these candidate genes in their ability to cause ET resistant growth, correlation of gene expression of each candidate gene was tested against *ESR1*/*PGR* expression in patient tumors from NeoAI (Figure S12). Partial correlation was observed between *ESR1*/*PGR* levels and *CETN2* (R=0.36/0.2), but not for other two genes. A ChIP-seq data set (47, 48) identified an *ESR1* binding peak close to the *CETN2* promoter but not for the other two candidates. Therefore, although ER-mediated regulation of some DDR genes cannot be ruled out as one mechanism underlying relationships with ET resistance, it is unlikely that it constitutes a common underlying mechanism for all the DDR candidate genes studies herein. Together, these data suggest that CEN- ER+ breast cancer cells, akin to MutL- cells, enable unchecked CDK4 activity, resulting in rapid G1/S transition even in the presence of ET.

To directly test whether inhibition of CDK4/6 can inhibit proliferation in CEN- ER+ breast cancer cells, MCF7 cells with stable knockdown of *CETN2*, *NEIL2* or *ERCC1* were exposed to the CDK4/6 inhibitors, palbociclib and abemaciclib. Control MCF7 cells demonstrated comparable sensitivity to both fulvestrant and CDK4/6 inhibitors, palbociclib (Figure 5F) and abemaciclib (Figure S11E), in keeping with published reports (12). However, downregulation of any one of the three candidate genes in MCF7 cells induced resistance to fulvestrant, but persistent sensitivity to both palbociclib (Figure 5F) and

abemaciclib (Figure S11E). These data provide preliminary support for a role for DDR-dysregulation in predicting ET resistance and sensitivity to CDK4/6 inhibitors.

Predictive value of candidate DNA damage repair gene dysregulation in ER+ breast cancer

To estimate the impact of DDR dysregulation as a novel class of ET resistance driver and a predictive marker for ET failure, the cumulative frequency of dysregulation, i.e. multiple or co-occurring downregulation of 3 of the 4 novel candidate genes discovered in this analysis, *CETN2*, *NEIL2*, *ERCC1,* mutation or copy number loss of the fourth candidate gene, *PRKDC*, and downregulation of the 2 previously known candidate genes, *MLH1* and *PMS2* was assessed in METABRIC (Figure 6A) and TCGA (Figure 6B). In both data sets, downregulation of one or a combination of these genes occurred in 40–60% of tumors from ER+ breast cancer patients who died within 5 years of diagnosis. A less significant enrichment for dysregulation of these genes was observed in ER+ breast cancer patients who died more than 5 years after diagnosis, suggesting that downregulation of these genes predisposes ER+ breast cancer to early ET failure consistent with intrinsic resistance.

To identify a DDR-low signature in ER+ breast cancer patients, a gene expression score was defined using mean normalized expression of each gene. The score was significantly lower in resistant tumors from NeoAI when compared against sensitive counterparts ($p=0.002$, Figure 6C). While this indicates that the score associates with ET resistance in patient tumors, the sensitivity of the score is ~70% and the specificity ~68%, indicating potential for further refinement of the signature by inclusion of other known factors, and mutational or copy number data.

Using this signature, the lowest and highest scoring quintiles of ER+ tumors were identified in METABRIC and Loi. The lowest scoring quintile associated with poor disease-specific and recurrence-free survival of patients with ER+ tumors in METABRIC ($p<0.001$, Figure 6D) and Loi ($p=0.09$, Figure 6E) indicating the feasibility of using this score to predict short-term outcomes in patient cohorts. Of note, this analysis also demonstrated better survival of patients in the upper quintile of the DDR signature score, suggesting dual validity of the score in predicting both worse and better response to ET.

DISCUSSION

This study presents a comprehensive characterization of the molecular landscape of canonical DNA repair pathway defects in ER+ breast cancer as it relates to response to ET. A previous epidemiological study examining a selected subset of BER proteins using immunohistochemistry identified XRCC1, APE1, SMUG1, and FEN1 as associating with ER+ breast cancer specific survival (49). However this study did not investigate a role for *NEIL2*, the only BER gene that was identified here. It is also noteworthy that the screening strategy outlined here did not identify *XRCC1* or *SMUG1* loss as associating with ET resistance, but this may be due to the stringent criteria required for a positive finding in our screening approach, i.e. independent validation in three data sets. We are unaware of other studies reporting a role for loss of any DDR candidates identified here in ET response. This is of specific interest with the publication of many large genome-wide studies of breast cancer, which were extremely valuable in our analyses. Although other studies using these

large data sets did not identify a role for DDR loss in endocrine resistance this is likely because in most data sets the diagnosis of endocrine resistance is based on relapse and death from disease, a phenotype that is highly dependent on the quality of the follow up. The use of neoadjuvant data sets in our analysis not only specifically addresses the intrinsic ET phenotype based on the highly prognostic nature of on-treatment Ki67 values, but also demonstrates that etiologocal diagnoses relating to endocrine resistance can be made very early on in the course of the disease, enabling interventions to address adverse biology early enough to improve overall outcomes.

The identification of DDR defects as regulators of response to ET also provide fundamental insights into into the etiology ER+ breast cancer. Previous studies have identified lower incidence of structural rearrangements in ER+ breast tumors when compared to either ER− or HER2+ tumors(50). Simultaneously, whole exome sequencing identified a subset of ER+ tumors with high somatic mutation load as associating with poor survival, whereas high mutation load in ER− tumors trended towards an association with better patient survival(42). The ability of ER+ breast cancer cells to grow in the presence of SSBR defects may reflect the evolutionary context of normal ER+ mammary cells, which are primed for sudden and rapid bursts of proliferation, associated with downregulation of many SSBR pathways(14). In contrast, ER+ mammary cells may find it more difficult to tolerate large genomic rearrangements, commonly associated with DSBR defects, as this is not part of their etiology. Further analysis of the unique role of NHEJ loss in endocrine treatment response is warranted.

In terms of alternative therapeutic strategies for CENMP- ER+ breast cancer patients, this study provides some preliminary but potentially important associations that warrant deeper investigation. MutL-defective, ET resistant, ER+ breast cancer cells and tumors are sensitive to CDK4/6 inhibitors(12), currently in clinical use in advanced disease settings. Preliminary functional investigations presented herein extend these observations, indicating that a common mechanism underlying endocrine resistance caused by disruption of multiple DDR candidate genes from different pathways can generate a disconnect between ER and CDK4/6 that is targetable with CDK4/6 inhibition.

The CEN score, which takes into account MMR, BER and NER pathway genes is a new starting point for distinguishing patients into those who are not likely to respond to ET and will require alternative treatments potentially including CDK4/6 inhibition. However, sophisticated algorithms and inclusion of additional DDR genes and other known factors that regulate ET response will be necessary to improve the sensitivity and specificity of this signature, particularly for the prediction of ET response. The ultimate valdation of our hypotheses awaits results from the many adjuvant CDK4/6 inhibitor trials that are ongoing.

In summary, the results of this study most clearly identify single-strand DNA damage repair defects as a novel class of ET resistance driver that may contribute to perhaps half of ER+ breast cancer patient deaths within the first 5 years after diagnosis. Detailed mechanistic studies focused on dysregulation of identified DDR components are ongoing to facilitate a better understanding of the fundamental connections between the ER, CDK4/6 and DNA

repair pathways in order to further refine the therapeutic approach that should be offered to these patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to acknowledge Mr. Jonathan T Lei, Drs. Shyam Kavuri and Eric Chang for scientific input.

Financial support: Research reported in this publication was primarily supported by Susan G. Komen Promise grant (PG12220321 to MJE), Cancer Prevention and Research Institute of Texas (CPRIT) Recruitment of Established Investigators award (RR140033 to MJE), Laura Ziskin award from Stand Up2 Cancer (MJE), and partially supported by Department of Defense Breakthrough Level 1 grant (BC170276 to SH). Clinical trial data accrual and analysis was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers U10CA180821 and U10CA180882 (to the Alliance for Clinical Trials in Oncology), U10CA077440 (legacy), U10CA180833, and U10CA180858. NeoPalAna trial was supported by Pfizer Pharmaceuticals and Susan G. Komen Promise Grant (MJE). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Davies C, et al. (2011) Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet (London, England) 378(9793):771–784.
2. Ma CX, Reinert T, Chmielewska I, & Ellis MJ (2015) Mechanisms of aromatase inhibitor resistance. Nature reviews. Cancer 15(5):261–275. [PubMed: 25907219]
3. Bose R, et al. (2013) Activating HER2 mutations in HER2 gene amplification negative breast cancer. Cancer discovery 3(2):224–237. [PubMed: 23220880]
4. Slamon DJ, et al. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science (New York, N.Y.) 235(4785):177–182.
5. Yersal O & Barutca S (2014) Biological subtypes of breast cancer: Prognostic and therapeutic implications. World Journal of Clinical Oncology 5(3):412–424. [PubMed: 25114856]
6. Goncalves R, Ma C, Luo J, Suman V, & Ellis MJ (2012) Use of neoadjuvant data to design adjuvant endocrine therapy trials for breast cancer. Nature reviews. Clinical oncology 9(4):223–229.
7. Cheang MCU, et al. (2009) Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. JNCI Journal of the National Cancer Institute 101(10):736–750. [PubMed: 19436038]
8. Barone I, et al. (2009) Expression of the K303R estrogen receptor-alpha breast cancer mutation induces resistance to an aromatase inhibitor via addiction to the PI3K/Akt kinase pathway. Cancer research 69(11):4724–4732. [PubMed: 19487288]
9. Osborne CK & Schiff R (2011) Mechanisms of endocrine resistance in breast cancer. Annual review of medicine 62:233–247.
10. de Groot AF, Kuijpers CJ, & Kroep JR (2017) CDK4/6 inhibition in early and metastatic breast cancer: A review. Cancer Treat Rev 60:130–138. [PubMed: 28961554]
11. Kwapisz D (2017) Cyclin-dependent kinase 4/6 inhibitors in breast cancer: palbociclib, ribociclib, and abemaciclib. Breast cancer research and treatment 166(1):41–54. [PubMed: 28741274]
12. Haricharan S, et al. (2017) Loss of MutL Disrupts CHK2-Dependent Cell-Cycle Control through CDK4/6 to Promote Intrinsic Endocrine Therapy Resistance in Primary Breast Cancer. Cancer discovery 7(10):1168–1183. [PubMed: 28801307]
13. Broustas CG & Lieberman HB (2014) DNA Damage Response Genes and the Development of Cancer Metastasis. Radiation research 181(2):111–130. [PubMed: 24397478]
14. Caldon CE (2014) Estrogen signaling and the DNA damage response in hormone dependent breast cancers. Frontiers in oncology 4:106. [PubMed: 24860786]

15. Nik-Zainal S, et al. (2012) Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell 149(5–10):979–993. [PubMed: 22608084]
16. Dietlein F, Thelen L, & Reinhardt HC (2014) Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. Trends in genetics : TIG 30(8):326–339. [PubMed: 25017190]
17. Nickoloff JA, Jones D, Lee SH, Williamson EA, & Hromas R (2017) Drugging the Cancers Addicted to DNA Repair. Journal of the National Cancer Institute 109(11).
18. Liberzon A, et al. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell systems 1(6):417–425. [PubMed: 26771021]
19. Liberzon A, et al. (2011) Molecular signatures database (MSigDB) 3.0. Bioinformatics (Oxford, England) 27(12):1739–1740.
20. Cline MS, et al. (Integration of biological networks and gene expression data using Cytoscape Cytoscape: a software environment for integrated models of biomolecular interaction networks. (1750–2799 (Electronic)).
21. Ellis MJ, et al. (2011) Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype--ACOSOG Z1031. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 29(17):2342–2349. [PubMed: 21555689]
22. Parker JS, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 27(8): 1160–1167. [PubMed: 19204204]
23. Zehir A, et al. (2017) Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 23(6):703–713. [PubMed: 28481359]
24. Olson JA Jr., et al. (2009) Improved surgical outcomes for breast cancer patients receiving neoadjuvant aromatase inhibitor therapy: results from a multicenter phase II trial. Journal of the American College of Surgeons 208(5):906–914; discussion 915–906. [PubMed: 19476859]
25. Liu X, Jian X, & Boerwinkle E (2011) dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 32(8):894–899. [PubMed: 21520341]
26. Ng PC & Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31(13):3812–3814. [PubMed: 12824425]
27. Hovelson DH, et al. (2015) Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. Neoplasia (New York, N.Y.) 17(4): 385–399.
28. Rhodes DR, et al. (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia (New York, N.Y.) 9(2):166–180.
29. Rhodes DR, et al. (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia (New York, N.Y.) 6(1):1–6.
30. Benjamini Y & Yekutieli D (2001) The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics 29(4):1165–1188.
31. Wang J, Vasaikar S, Shi Z, Greer M, & Zhang B (2017) WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic acids research.
32. Haricharan S & Brown P (2015) TLR4 has a TP53-dependent dual role in regulating breast cancer cell growth. Proceedings of the National Academy of Sciences of the United States of America 112(25):E3216–3225. [PubMed: 26063617]
33. Ellis MJ, et al. (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 486(7403):353–360. [PubMed: 22722193]
34. Ellis MJ, et al. (2017) Ki67 Proliferation Index as a Tool for Chemotherapy Decisions During and After Neoadjuvant Aromatase Inhibitor Treatment of Breast Cancer: Results From the American College of Surgeons Oncology Group Z1031 Trial (Alliance). Journal of clinical oncology : official journal of the American Society of Clinical Oncology 35(10):1061–1069. [PubMed: 28045625]

35. Urruticoechea A, Smith IE, & Dowsett M (2005) Proliferation marker Ki-67 in early breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 23(28):7212–7220. [PubMed: 16192605]
36. Liu J, et al. (2016) GATA3 mRNA expression, but not mutation, associates with longer progression-free survival in ER-positive breast cancer patients treated with first-line tamoxifen for recurrent disease. Cancer Lett 376(1):104–109. [PubMed: 27018307]
37. Chimge NO, et al. (2016) RUNX1 prevents oestrogen-mediated AXIN1 suppression and beta-catenin activation in ER-positive breast cancer. Nat Commun 7:10751. [PubMed: 26916619]
38. Thewes V, et al. (2015) Reprogramming of the ERRalpha and ERalpha target gene landscape triggers tamoxifen resistance in breast cancer. Cancer research 75(4):720–731. [PubMed: 25643697]
39. Iwamoto T, et al. (2012) Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 30(7):729–734. [PubMed: 22291085]
40. Loi S, et al. (2008) Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC genomics 9:239. [PubMed: 18498629]
41. Griffith OL, et al. (2017) The prognostic effects of somatic mutations in ER-positive breast cancer. bioRxiv.
42. Haricharan S, Bainbridge MN, Scheet P, & Brown PH (2014) Somatic mutation load of estrogen receptor-positive breast tumors predicts overall survival: an analysis of genome sequence data. Breast cancer research and treatment 146(1):211–220. [PubMed: 24839032]
43. Bruna A, et al. (2016) A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. Cell 167(1):260–274.e222. [PubMed: 27641504]
44. Mjelle R, et al. (2015) Cell cycle regulation of human DNA repair and chromatin remodeling genes. DNA Repair (Amst) 30:53–67. [PubMed: 25881042]
45. Nalepa G & Clapp DW (2018) Fanconi anaemia and cancer: an intricate relationship. Nature reviews. Cancer 18(3):168–185. [PubMed: 29376519]
46. Murray JM & Carr AM (2018) Integrating DNA damage repair with the cell cycle. Curr Opin Cell Biol 52:120–125. [PubMed: 29587168]
47. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, & Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat Genet 43(1):27–33. [PubMed: 21151129]
48. Joseph R, et al. (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. Mol Syst Biol 6:456. [PubMed: 21179027]
49. Abdel-Fatah TM, et al. (2014) Is there a role for base excision repair in estrogen/estrogen receptor-driven breast cancers? Antioxidants & redox signaling 21(16):2262–2268. [PubMed: 25111287]
50. Kwei KA, Kung Y, Salari K, Holcomb IN, & Pollack JR (2010) Genomic instability in breast cancer: pathogenesis and clinical implications. Molecular oncology 4(3):255–266. [PubMed: 20434415]

Statement of Translational Relevance

Estrogen receptor positive (ER+) breast cancer is treatable with endocrine drugs that interrupt estrogen receptor function, but fatal drug resistance occurs in at least 30% of cases. A comprehensive analysis of the molecular landscape of DNA repair defects in ER + breast cancer patient tumors reported here identifies defects in select DNA repair pathways as a novel class of endocrine treatment resistance driver occurring in ~40% of endocrine treatment resistant ER+ breast cancer patients. Candidate DNA damage repair genes identified are experimentally shown to be linked by common upregulation at the G1/S transition, suggesting that loss of DNA proof reading pathways disrupts ER regulation of the cell cycle, and therefore, response to endocrine treatment. A combined DDR signature score was developed that predicted poor outcome in multiple patient cohorts, which will have immense translational implications in stratifying patients who would not respond to endocrine therapy but will be good candidates for CDK4/6 inhibitors based treatment. These findings, therefore, significantly increase understanding of factors underlying response to standard-of-care in ER+ breast cancer patients and identify an alternative targeted therapeutic strategy for this subset of patients.

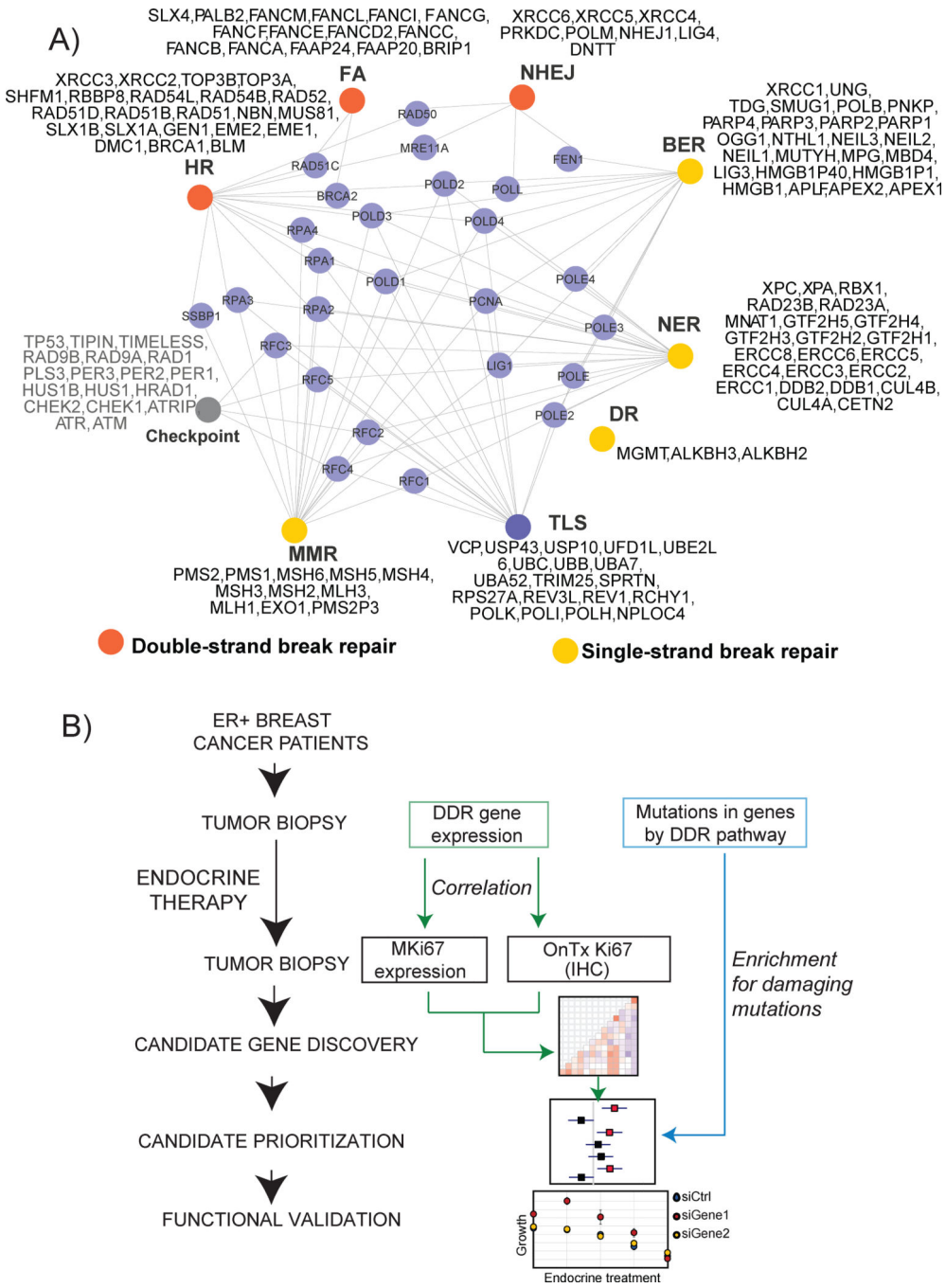


Figure 1. Study outline.

A) Network view of different DDR pathways along with the shared genes (gray nodes) and unique genes (indicated by name, adjacent to pathway name). Pathways associated with SSBR are denoted as yellow nodes and DSBR denoted as orange nodes. Lines indicate pathways that share common genes. Mismatch repair (MMR), nucleotide excision repair (NER), base excision repair (BER), non-homologous end joining (NHEJ), homologous recombination (HR), Fanconi Anemia (FA), trans-lesion synthesis (TLS) and direct repair (DR). B) Schematic representation of design of screening approach to identify DDR

pathways and genes associated with ET response. Grey boxes indicate data accrued at baseline and cyan boxes indicate data accrued on-AI treatment. Supporting data with detailed schema is presented in Figure S1.

A)

Genes	Correlation FDR Ki67 (mRNA)	FDR Ki67 (IHC)	DDR function Pathways
CETN2	0.10	0.04	NER
CUL4A	0.92	0.11	NER
ERCC1	0.53	0.02	NER
FANCF	0.02	0.43	FA
GTF2H5	0.16	0.34	NER
MLH1	0.12	0.00	MMR
MLH3	0.03	0.03	MMR
MNAT1	0.36	0.05	NER
MPG	0.12	0.09	BER
MSH3	0.07	0.04	MMR
NEIL2	0.12	0.32	BER
PMS2	0.06	0.43	MMR
XPA	0.02	0.03	NER
ESR1	0.00	0.05	Non-DDR Genes implicated in ET resistance (Iwamoto, 2012, JCO)
GATA3	0.11	0.95	Non-DDR Genes implicated in ET resistance (Liu, 2016, Cancer Lett.)
RUNX1	0.08	0.18	Non-DDR Genes implicated in ET resistance (Thewes, 2015, Cancer Res.)

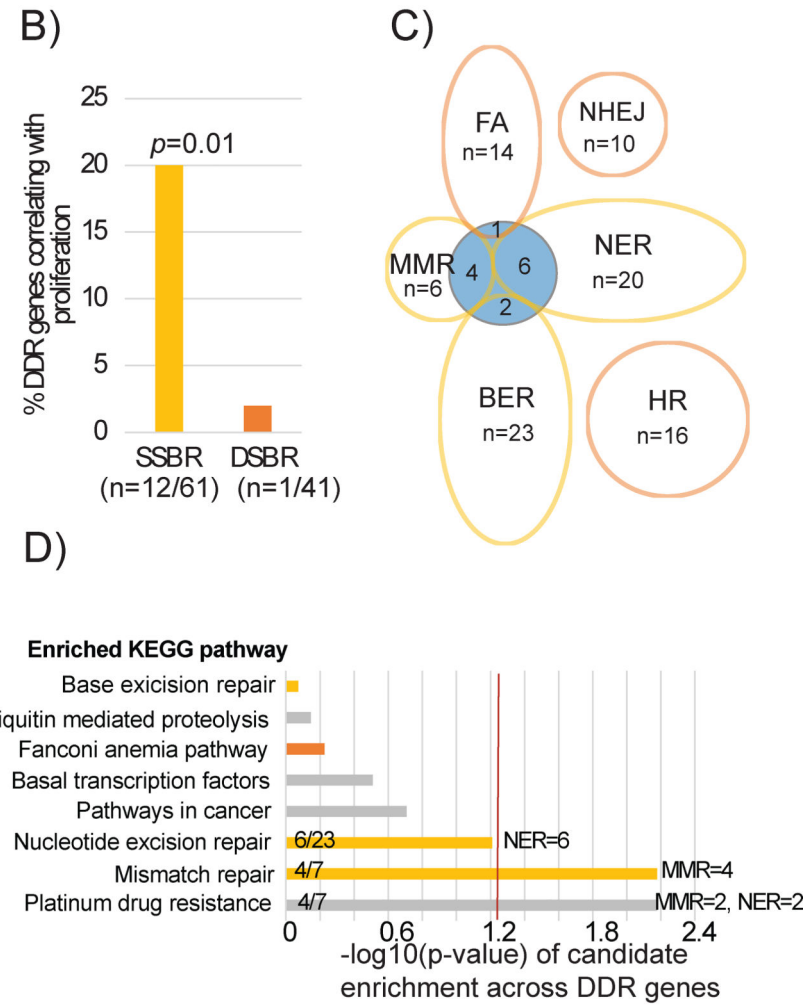


Figure 2. RNA levels of MMR, BER and NER genes associate inversely with on-endocrine treatment Ki67.

A) Table describing 13 candidate genes with significant correlation between RNA levels and on-treatment Ki67 in ER+ tumors from NeoAI. Pearson correlation analysis was used to determine the correlation coefficient. False discovery rate (FDR) is denoted in the table for specific correlations. Blue boxes indicate correlations where FDR<20%. Pathways to which candidate genes belong are noted in the DDR function column. Yellow boxes indicate SSBR pathways and orange boxes, DSBR. As a positive control for the analysis, three genes previously implicated in response to ET: *ESR1*, *GATA3* and *RUNX1*, are included. B) Bar graph indicating enrichment for SSBR genes in the list of 13 candidate genes. Fisher's exact test determined p-value. C) Venn diagram depicting proportion of genes from each DDR pathway that was implicated in ET resistance from correlation analysis in A. Blue circle indicates candidate gene population. D) KEGG pathway enrichment analysis of the candidate gene list against all DDR genes used in the analysis revealed significant enrichment of indicated pathways. Number of genes from candidate list contributing to each enriched pathway is listed along the bars. A –log10 (p-value) of 1.2 denotes a $p<0.05$.

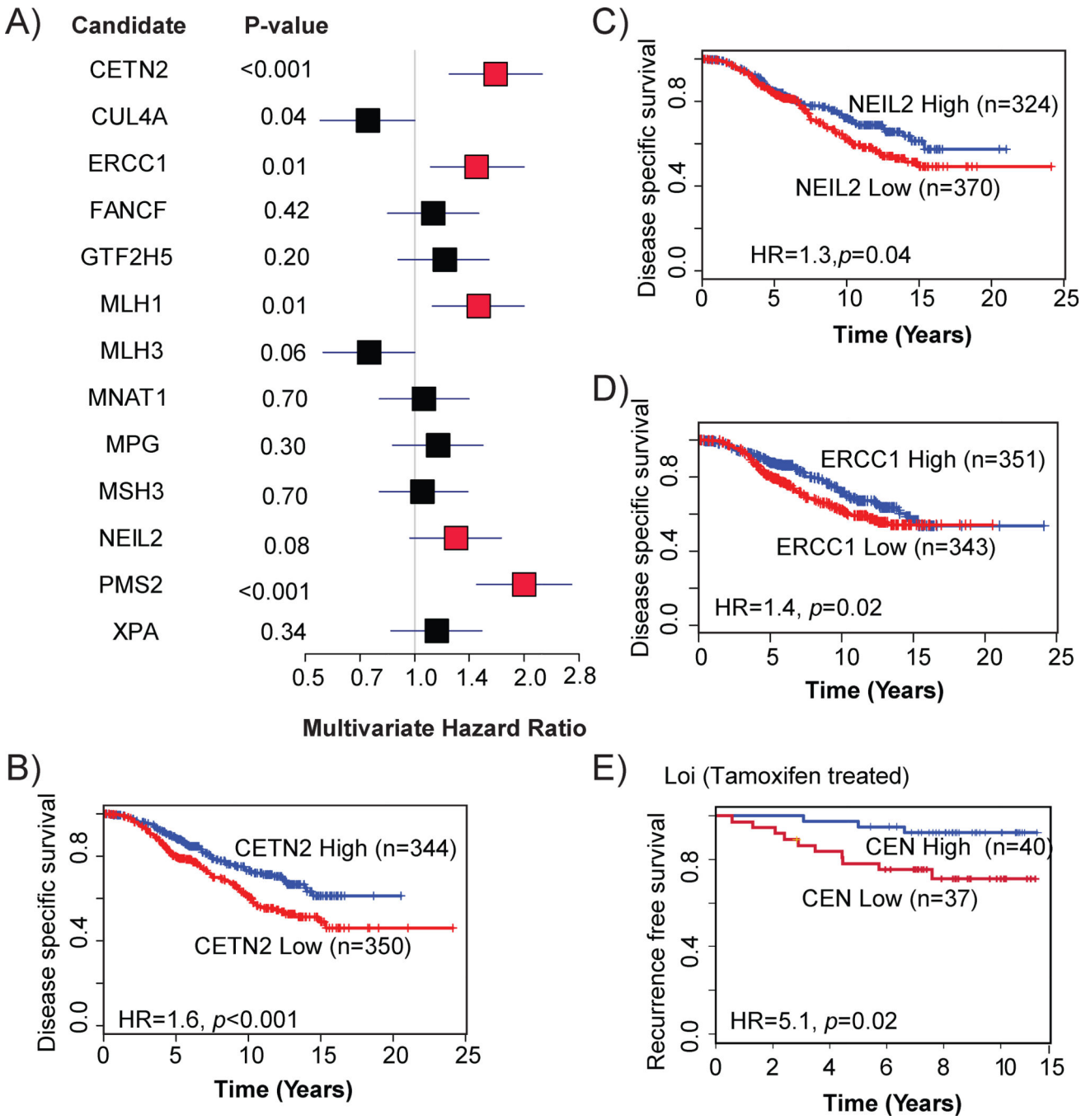


Figure 3. *CETN2*, *NEIL2* and *ERCC1* loss associates with poor survival of ER+ breast cancer patients.

A) Forest plot summarizing results of multivariate analysis of the 13 candidate genes in METABRIC. Other factors included in the analysis were tumor size, grade and node positivity. Boxes denote hazard ratio (HR) based on overall survival outcome, and error bars the 95% confidence interval. HR for genes whose dysregulation associated with poor survival ($p \leq 0.05$) by univariate analysis (presented in Figure S2) are shown as red boxes. B-D) Kaplan-Meier curves depicting disease specific survival of patients with luminal breast cancer treated with ET whose tumors have low (mean-1.5 standard deviation) *CETN2* (B),

NEIL2 (C) and *ERCC1* (D) expression (red) in METABRIC data set. Kaplan-Meier curves for HER2-enriched and basal-like tumors are presented in Figure S3. (E) Kaplan-Meier curves depicting recurrence free survival of tamoxifen treated ER+ breast cancer patients whose tumors had low expression of *CETN2*, *ERCC1* and *NEIL2* (CEN Low in red) in Loi data set. Individual Kaplan-Meier curves presented in Figure S4. All HRs were calculated using Cox Regression and log-rank p-value determined significance of differences in survival.

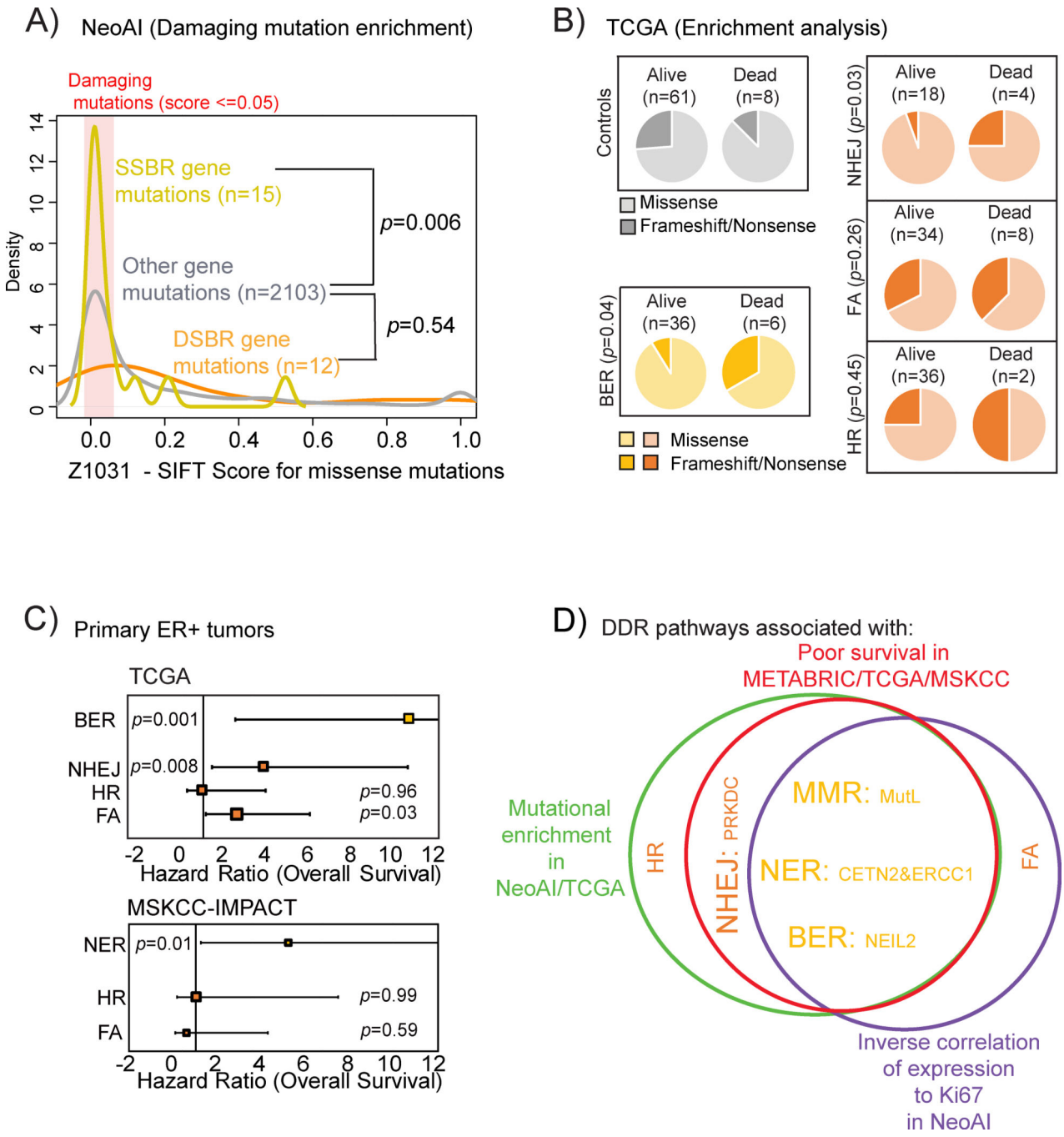


Figure 4. NER, BER and NHEJ genes are enriched for damaging mutations in endocrine treatment resistant tumors.

A) Enrichment analysis for prevalence of predicted damaging mutations (based on SIFT scores: lower the SIFT score, the more damaging the mutation is predicted to be) in SSBR and DSBR pathways compared to genome-wide prevalence in tumors from NeoAI. Significant p-values were determined by Wilcoxon test analysis. Similar analysis for each individual DDR pathway is presented in Figure S5. B) Pie charts comparing proportion of missense (light yellow – SSBR, light orange - DSBR) and frameshift/nonsense (yellow – SSBR, orange - DSBR) mutations in SSBR and DSBR genes relative to proportion in

control gene set (grey). Z-statistic for two population proportions was used to determine significant differences in proportion of missense to frameshift/nonsense mutations in patients who remained alive to maintain adequate sample size for the test. C) Forest plots depicting hazard ratios for overall survival of patients from TCGA (above) and MSKCC-IMPACT (below) with ER+ tumors harboring non-synonymous mutations in indicated pathways. Log rank test was used to determine significance and Cox Regression Proportional Hazards generated univariate hazard ratios. Supporting data investigating a role for NHEJ gene mutation in ER+ breast cancer survival is presented in Figure S6, and analyses controlling for replication defects, genome instability and mutation load are presented in Figures S7–8. D) Venn diagram and word cloud (yellow text, SSBR and orange text, DSBR) summarizing candidate pathways that significantly associate with poor survival of ER+ breast cancer patients (red) based on mutational (green) or transcriptomic (violet) dysregulation. MMR, NER and BER pathways are identified at the intersection of all analyses. Larger font size indicates greater confidence.

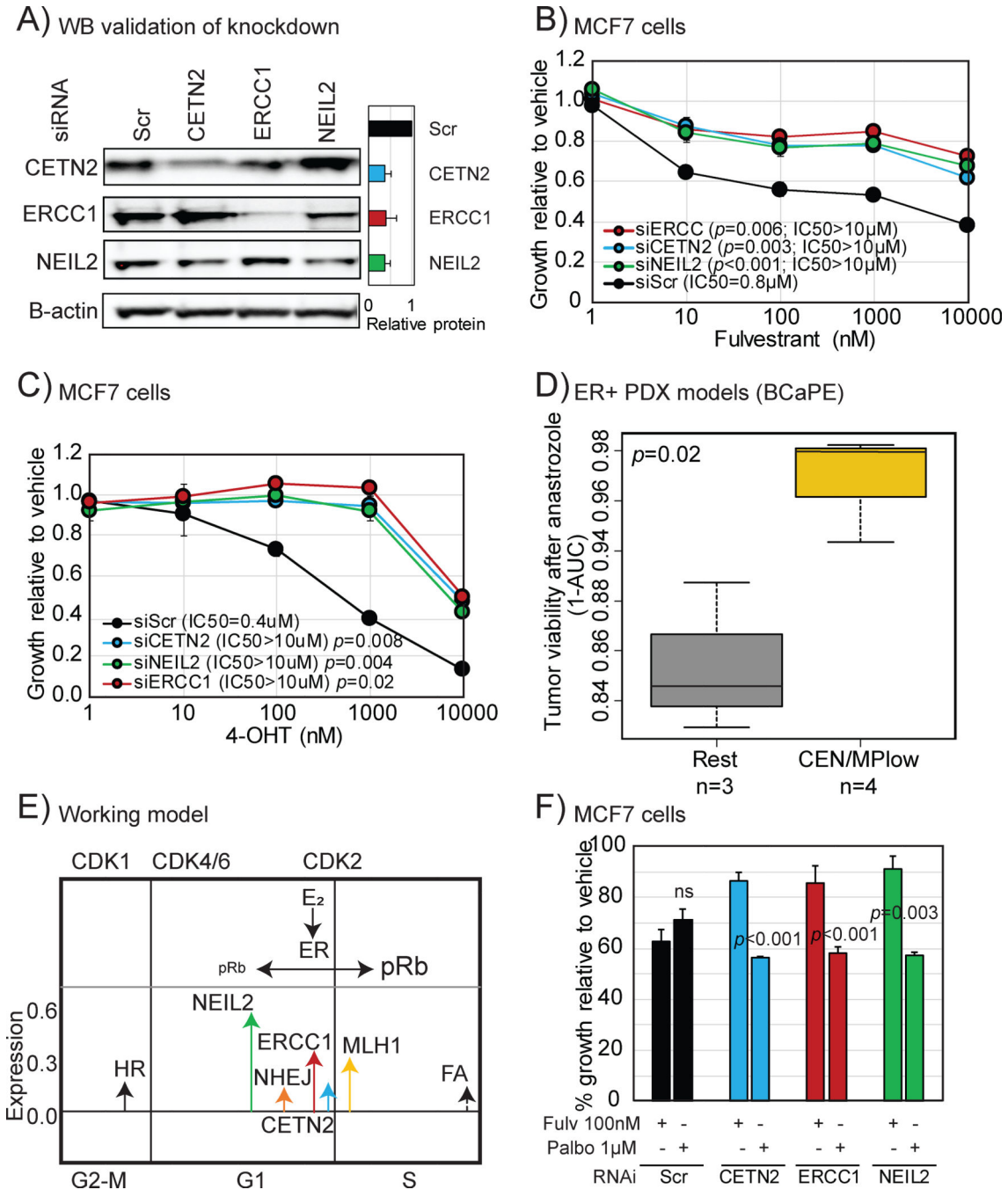
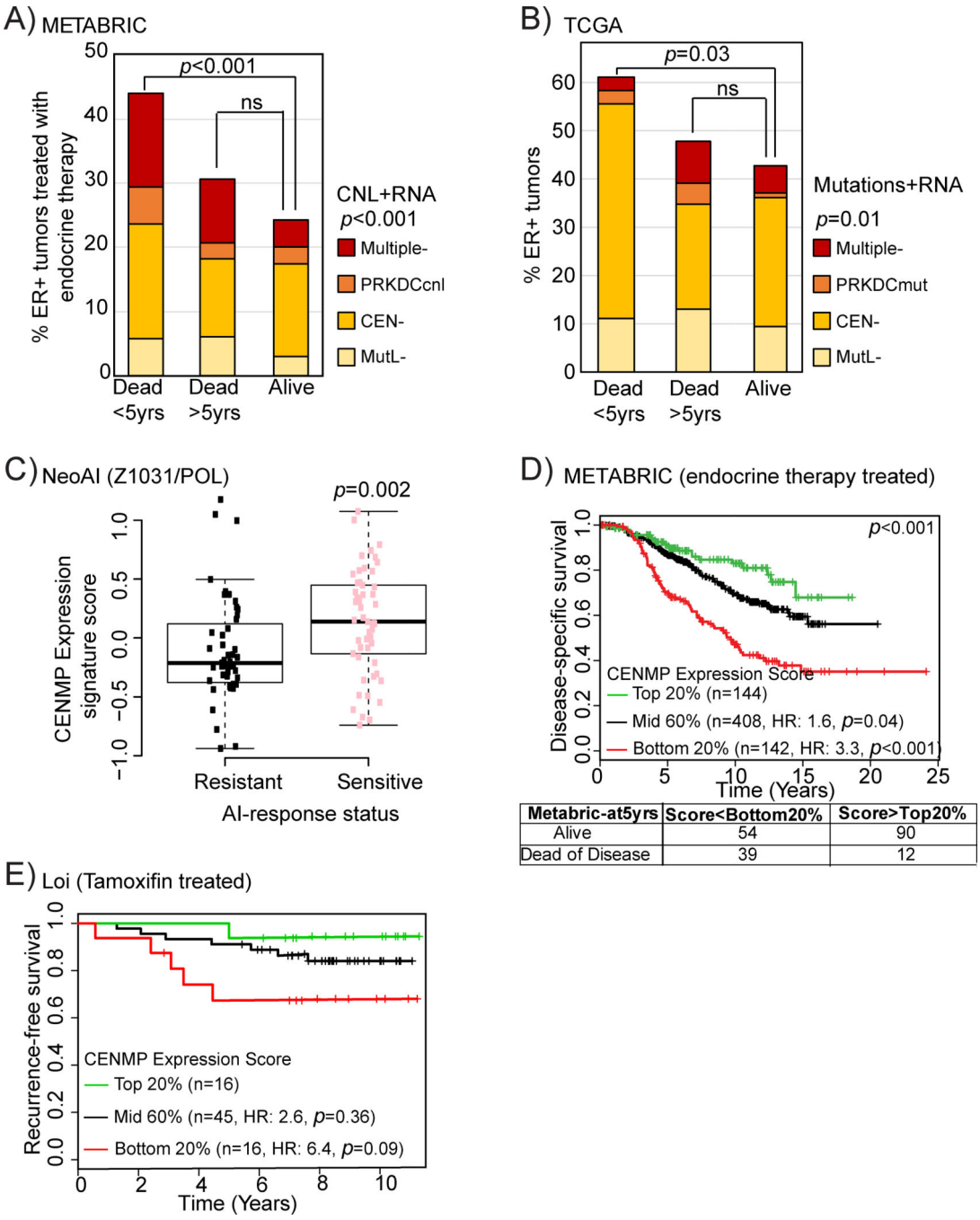


Figure 5. Inhibition of CETN2, NEIL2 and ERCC1 induces resistance to all classes of endocrine therapy in ER+ breast cancer cells and PDXs.

(A) Western blot validation of siRNA-mediated knockdown of *CETN2*, *NEIL2* and *ERCC1* respectively in MCF7 cells. Results from three independent experiments are depicted. Columns represent the mean and error bars the standard deviation. RNA level validation of knockdown is presented in Figure S9A. (B-D) Dose response curves of MCF7 cells with transient inhibition of *CETN2*, *NEIL2* or *ERCC1* treated with fulvestrant (B) or 4-hydroxy-tamoxifen (C). Dose response to estrogen stimulation is presented in Figure S9B. IC50 values were calculated from three independent dose curves for each condition and Student's

t-test used to determine significant differences in IC50 values. nM, nanomolar. Independent validation in a second cell line is presented in Figures S9C–F, and orthogonal validation of knockdown results are presented in Figure S10. (D) Box plot depicting tumor viability *in vivo* after anastrozole treatment of 7 ER+ PDX lines from BCaPE, calculated using area under the curve (AUC) measurements. CEN: CETN2, ERCC1, NEIL2; MP: MLH1, PMS2. Wilcoxon Rank Sum test determined p-value. (E) Working model indicating peak expression levels of *NEIL2*, *ERCC1*, *MLH1*, and *CETN2* genes across the cell cycle. Data generated from two independent double thymidine block experiments (www.dnarepairgenes.com). Cumulative peak expression level of all genes in NHEJ, HR and FA pathways also indicated. Y-axis indicates relative gene expression level and X-axis is plotted based on number of hours post release of double thymidine block. Implication of CDKs and estrogen stimulation in the cell cycle is based on published reports. Supporting mechanistic data is presented in Figures S11–12. (F) Bar graphs represent growth inhibition, relative to vehicle treated cells, in response to 100nM of fulvestrant or 1 μM of Palbociclib, CDK4/6 inhibitor in MCF7 cells stably expressing pooled RNAi oligos against *CETN2*, *ERCC1*, *NEIL2* or scrambled control. Student's t-test determined p-values by comparing growth inhibition in response to Palbociclib against that in response to fulvestrant.



Metabric-at5yrs	Score<Bottom20%	Score>Top20%
Alive	54	90
Dead of Disease	39	12

Figure 6. Cumulative incidence and predictive potential of *CETN2*, *NEIL2*, *ERCC1*, *MLH1* and *PMS2* (CENMP) deficiency.

A-B) Stacked columns indicating cumulative frequency of dysregulation (mutation or underexpression) of *CETN2*, *ERCC1*, *NEIL2* (CEN-); *MLH1*, *PMS2* (MutL-); and *PRKDC* mutation (mut) or copy number loss (cnl) in ER+ breast tumors from METABRIC (A) and TCGA (B). Fisher's exact test determined p-values. C) Box plots describing CENMP expression signature score in tumors from patients based on their response to AI-treatment. Wilcoxon Rank Sum test determined p-values. D-E) Kaplan-Meier survival curves evaluating separation based on CENMP score in ET treated ER+ patients from METABRIC

(D) and Loi (E) data sets. Cox Regression identified hazard ratio (HR) and log rank test determined p-values for survival analyses.