Cell PRESS
Open ACCESS

EXTENDED EXPERIMENTAL PROCEDURES

Breast Cancer Specimen Acquisition and Processing of Cells for Engraftment

All human tissues for these experiments were processed in compliance with NIH regulations and institutional guidelines, approved by the Institutional Review Board at Washington University. Tumors from breast cancer patients were obtained via core needle biopsy, skin punch biopsy, or surgical resection after informed consent. Fresh tumor tissues from each patient were divided into three parts. One part was immediately placed into a frozen section cassette filled with Optimum Cold Temperature Medium (OCT, Tissue-Tek; Miles, Elkhart, IN) and snap frozen in liquid nitrogen for DNA and RNA analysis, one was fixed in 10% buffered formalin and embedded into paraffin blocks, and one for engraftment was placed in ice-chilled DMEM/F12 medium and rapidly processed for engraftment. After removal of necrotic tissue and fat, tumor specimens for generating xenografts were chopped with scalpel blades to yield 1 × 1 mm pieces under aseptic sterile conditions. For the HIM technique (Kuperwasser et al., 2004) tissue fragments were gently dissociated on a rotary shaker at 37°C for 3-4 hr in DMEM medium supplemented with 5ug/ml insulin, 0.5ug/ml hydrocortisone, 10ng/ml cholera toxin, 2% bovine serum albumin, 0.3mg (300U)/ml hyaluronidase (all from Sigma), 1 × antibiotic-antimycotic (GIBCO), and3mg/ml collagenase A (Roche). After dissociation, the resulting cells were washed twice with PBS, trypsinized for 5 min and then filtered through a 45-μm cell strainer. The single cell suspension was either injected into mice or cryopreserved.

Animals, Surgeries, and Tumor Xenografts

Three-week-old NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NOD/SCID) female mice were obtained from NCI and Jackson Labs. All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee at Washington University in St. Louis. The animals were housed under controlled temperature; humidity and lighting conditions in filter top cages. Surgeries were performed under sterile conditions. For the human-in-mouse technique, mouse mammary fat pads were humanized as previously described. Briefly, three-week-old female mice were anesthetized for removing the mammary epithelium from the #4 inguinal mammary fat pad. Three to four weeks post clearance, 2.5 × 10^5 irradiated and 2.5 × 10^5 unirradiated EG fibroblasts (Gift from C Kuperwasser) (5 Gy) were injected into the cleared fat pads using a 27-gauge needle. Four to six weeks after the humanization, 1 × 10^6 cells dissociated from primary tumors were mixed with 0.5 × 10^5 EG fibroblasts and resuspended in 40μl of a 1:1 Matrigel(BD Biosciences): Collagen (Upstate Biotechnology, Lake Placid, NY) mixture and injected into the humanized mammary glands. When xenograft tumors reached 1.5 cm in diameter they were harvested and divided into three parts. One part was immediately placed into a frozen section cassette filled with OCT and snap frozen in liquid nitrogen for DNA and RNA analysis, one was fixed in 10% buffered formalin, and one was dissociated into single cell suspension for serial transplantation. For later experiments using direct implantation of human tumor tissues into mammary fat pads, human tissues were chopped into 1-2mm pieces under a dissecting microscope and implanted into mammary fat pad of NOD/SCID mice. PDX models can be accessed through an enquiry to the HAMLET core web site at http://digitalcommons.wustl.edu/hamlet/.

Western Blot Analysis

Frozen WHIM tumor samples were extracted for soluble proteins using PULSE Tubes (FT500-ND, Pressure BioSciences) in RIPA lysis buffer (20mM Tris-HCl, pH 7.5, 150mM NaCl, 10% glycerol, 1% NP-40, 0.1% SDS, 0.25% sodium deoxycholate, protease inhibitors (EDTA-free Complete Mini, Roche), 1 mM EDTA, 1 mM PMSF and phosphatase inhibitors (1 mM NaPPi, 1 mM Na_3VO_4, 10 mM NaF)). Soluble proteins from sub-confluent breast cancer cell lines were extracted with RIPA lysis buffer for 10 min at 4°C. Tumor and cell line lysates were clarified by centrifugation for 5mins at 13,000 rpm, measured for protein concentration by Bradford assay, and mixed with SDS sample buffer for immunoblotting. Proteins (30μg) were separated on 4%–12% gradient or 10% Bis-Tris gels (Invitrogen), transferred to nitrocellulose membrane, blocked with 5% skim milk in TBS/0.1% Tween-20 (TBST) for 1hr at room temperature and incubated overnight at 4°C with primary antibodies. The following primary antibodies were used: rabbit polyclonal N-terminal ERα antibody (Ab75635, Abcam, 1:1000), rabbit monoclonal C-terminal ERα antibody (Rm-9101-S1, Thermo, 1:1000), rabbit monoclonal YAP1 antibody (Ab52771, Abcam, 1:5000), rabbit monoclonal progesterone (PR) antibody (#8757, Cell Signaling Technology, 1:1000), mouse monoclonal CD44 antibody (#3570, Cell Signaling Technology, 1:1000), polyclonal rabbit HER2 antibody (A0485, Dako, 1:1000), mouse monoclonal cytokeratin 5/6 antibody (MAB1620, Millipore, 1:500), rabbit polyclonal DYKDDDDK (Flag epitope) Tag antibody (#2368, Cell Signaling Technology, 1:1000), rabbit monoclonal TFF1 antibody (#2801-1, Epitomics, 1:1000) and goat polyclonal actin (SC-1616, Santa Cruz, 1:5000). After washing with TBST, membranes were incubated with appropriate HRP-conjugated secondary antibodies for 1hr at room temperature, followed by washing, and developed using a chemiluminescence substrate (SuperSignal West Pico, Thermo) and visualization on a ChemiDoc Imaging System (Bio-Rad).

Gene-Expression Analysis 4X44K Agilent Arrays

Gene expression microarray processing, data quality control and processing, and PAM50 subtype classification have been previously described (Parker et al., 2009). Unsupervised hierarchical clustering was conducted using originating tumor and WHIM sample paired samples using all genes of the microarrays except the stromal-related which were identified after performing a two-class

paired Significance Analysis of Microarrays (SAM) with a false discovery rate of 0% between 18 paired progenitor human tumors and xenografts. A list of 973 genes was found to be significantly upregulated in human tumors compared to xenografts (here called stromal-related genes). The complete list of up- and down-regulated genes can be found in Table S2B. All microarray data have been deposited in the Gene Expression Omnibus (GEO) under the accession number GEO: GSM41685.

Gene-Expression Analysis 244K Agilent Custom Arrays

Breast samples and WHIMs were profiled as described previously (Hu et al., 2006) using 244K human oligo microarrays (Agilent Technologies, Santa Clara, CA, USA). The probes or genes for all analyses were filtered by requiring the lowest normalized intensity values in both sample and control to be > 10. The normalized log2 ratios (Cy5 sample/Cy3 control) of probes mapping to the same gene (Entrez ID as defined by the manufacturer) were averaged to generate independent expression estimates. For Cy3-controls, we used Stratagene Human Universal Reference enriched with equal amounts of RNA from the MCF7 and ME16C cell lines. Genes were median-centered and samples were standardized to zero mean and unit variance. All microarray data are available in the University of North Carolina (UNC) Microarray Database (https://genome.unc.edu/) and have been deposited in the Gene Expression Omnibus (GEO) under the accession number GEO: GSE46604.

Reverse Phase Protein Array

Cell lysates were obtained by homogenizing tumor tissue with lysis buffer and then serially diluted for quantification by a Tecan liquid handling robot (Tecan Systems, Inc., San Jose, CA). The lysates were mixed with sodium dodecyl sulfate (SDS) boiled and printed onto nitrocellulose-coated glass slides (FAST Slides, Schleicher & Schuell BioScience, Inc., Keene, NH) with an automated Aushon arrayer (Aushon Biosystems, Burlington, MA). For 5-step dilution, as many as 1056 samples can be printed on a single slide. Six pooled cell line controls were also printed in 5-step dilution series to provide accurate quantification and quality control. The 3,3′-diaminobenzidine tetrachloride (DAB)-based DAKO signal amplification system (DAKO, Copenhagen, Denmark) was used to detect and amplify antibody-binding intensity, and a ~~-A~~ biotinylated secondary antibody was used as a starting point for signal amplification. Signal intensity was measured by scanning the slides and quantifying with MicroVigene software (VigeneTech Inc., Carlisle, MA) The protein concentration levels in the samples were estimated using Supercurve (version 1.5.0) developed by the Department of Bioinfomatics and Computational Biology at the University of Texas MD Anderson Cancer Center, http://bioinformatics.mdanderson.org/OOMPA). The program fits a single nondecreasing spline curve using all the dilution series on a slide with the signal intensity as the response variable and the dilution steps as independent variable (Wang et al., 2010) to estimate some basic parameters, and then estimate the IC_{50} of each dilution series. Before any downstream analysis was carried out, the data were normalized to correct the loading effect. The loading correction procedure can be described as followings: first median-center each marker (on log2 scale, across all samples), and then the median is calculated for each sample (across all markers). Finally, the sample medians above are subtracted from the raw data of each sample.

Whole-Genome Sequencing and Capture Validation

Thirteen patients with blood, tumor, and xenograft were selected for whole genome sequencing (WGS). Detailed histories for these patients and xenografts are provided in Table S2A. Libraries were prepared using unamplified genomic DNA from blood (normal), tumor, and xenograft samples. Paired-end sequencing was performed on the Illumina platform, as previously described (Walter et al., 2012; Ellis et al., 2012). Variant calling and validation of all mutations using liquid hybridization capture were performed, as previously described (Welch et al., 2012).

RNA Sequencing

mRNA-seq was performed as previously described (Cancer Genome Atlas Research Network, 2012b); ~1ug of total RNA was used with the Illumina Tru-Seq kit to create mRNA-seq libraries; these libraries were sequenced as one library/sample per lane, with a 2x50bp sequencing configuration. The reads were then mapped using the MapSplice algorithm (Wang et al., 2010). The UNC analysis pipeline was used to generate the .BAM files. Expressed gene fusions were nominated using ChimeraScan v0.4.3 (Iyer et al., 2011) with default parameters. Gene fusion nominations were required to have two independent spanning junction reads.

Quantitative Real-Time PCR to Determine ESR1 Amplification

Primers were designed using Primer3 (Rozen and Skaletsky, 2000). Assays were optimized and run on a Viaa 7 Real-Time PCR System (Life Technologies) using 10ng of genomic DNA for each sample according to the manufacturer's instructions. One set of primers was used for each of the control genes FAM38B and ASXL2, as described previously (Reis-Filho et al., 2008). Three primer sets were used to interrogate the amplification at locations 1, 2, and 3 (as specified in Figure 5B). p values were computed by a one-way ANOVA using Dunnett's post hoc test. Three replicates were performed for each qPCR.

Tests for Estradiol Dependence on ER+ WHIM Lines

2-5 × 10^6 WHIM cells were injected into *ovariectomized (OVX) NOD/SCID mice* (Jackson Laboratories). Treatment cohorts were grouped as OVX and OVX + 17β-estradiol (E2) pellets by implantation of 90-day release E2 pellets (1.7 mg/pellet; Innovative Research

of America). The E2 pellets were implanted either a few days prior to surgery or when the tumors reached 3-4mm in diameter. For treatment of WHIM18 with fulvestrant, female CB.17 SCID mice from the University of Michigan breeding colony were used. 4-5 week-old mice were implanted s.c. with a 2-3mm tumor piece from an actively growing WHIM18 tumor. Tumors were allowed to grow for 60 days, until tumors were 150-400mm^3, at which time they were size matched and randomized into treatment groups with 8 mice per group. Mice were treated weekly with either Vehicle or 250mg/kg fulvestrant (AstraZeneca) by subcutaneous injection for 4 weeks. Tumors were measured in two dimensions and tumor size was calculated as Tumor volume (mm^3) = (A X B^2)/2 where A and B are the tumor length and width (in mm), respectively.

Cloning and Lentiviral Transduction of ESR1 and ESR1-YAP1 into Cell Lines

Wild-type ESR1, the Y537S or Y537N ESR1 mutants and the ESR1-YAP1 fusion were fused to a FLAG tag at their C-termini and cloned into the lentiviral vector pFLRu-FH (Feng et al., 2010). Y537S and Y537N mutations were first introduced into ESR1 using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) with an ESR1-encoding plasmid (Accession number NM_000125.1, GeneCopoeia Inc.) as the template. A FLAG-tag (DYKDDDDK) was then added to the C terminus of full-length wild-type and mutant (Y537S and Y537N) ESR1 by PCR amplification, followed by shuttling into the designation lentiviral vector pFLRu-FH. The ESR1/YAP1 fusion cDNA was created by overlapping PCR extension/amplification to make a fusion gene product encoding ESR1 exons 1–4 (amino acids 1-365) in-frame with YAP1 exons 4–9 (amino acids 230–504). ESR1 and YAP1 (Accession number BC038235.1, Open Biosystems) cDNAs were used as templates in PCR reactions to create overlapping 5′ and 3′ fragments of the ESR1/YAP1 fusion gene, respectively. These cDNA fragments were used as templates in a second PCR reaction to create the full length ESR1/YAP1 cDNA containing the FLAG tag at the C terminus, followed by shuttling into the pFLRu-FH lentiviral vector. Primer sequences for cloning are available upon request. To make lentiviral particles, pFLRu-FH vector DNAs (encoding YFP, ESR1(wt), ESR1(Y537S), ESR1(Y537N) and ESR1/YAP1) were cotransfected with the packaging plasmids into HEK293T cells using Fugene 6 (Roche). 48 hr posttransfection, culture media containing different viruses were added to T47D and MCF7 cells in the presence of polybrene followed by 3-day puromycin selection for stable expression. Transgene expression was verified by western blot analysis for wild-type and mutant ER mutant proteins and for ESR1-YAP1.

Tissue Culture and In Vitro Growth Assays

HEK293T cells to generate lentivirus (ATCC) were grown in DMEM medium (GIBCO) with 10% fetal bovine serum (FBS) and 1X Pen/Strep/Fungizone Mix (GIBCO). T47D, MCF7 and MDA-MB-231 cells obtained from ATCC were propagated in RPM1 1640 containing 10% fetal bovine serum (FBS) with antibiotic and supplements (50 μg/mL gentamycin, pyruvate, 10 mM HEPES and glucose to 4.5 g/L) in a humidified 37°C incubator containing 5% CO2. For studies involving endocrine treatments with short-term estrogen deprivation, derivatives of T47D and MCF7 cells were maintained in phenol red-free RPMI 1640 containing 10% charcoal stripped FBS, antibiotics and supplements (CSS medium) for approximately 1–3 weeks prior to experimental treatments. The MCF7 LTED cell line, created by estrogen deprivation of MCF7 cells for approximately 2 years (Sanchez et al., 2011), was grown in CSS medium. For growth assays, cells cultured in CSS medium were plated in 96-well dishes. The day after plating, the Day 0 plate was read to obtain baseline values by measuring Alamar Blue reduction (540λ_{Ex}/ 590λ_{Em}) with a fluorescent microplate reader (Synergy H1 reader, BioTek Instruments). Treatments were then done in quadruplicate without or with fulvestrant (500 nmol/L, Sigma) and estradiol (E_2, 10 nmol/L, Sigma). Medium was replenished every 3-4 d and cell growth was assessed at the indicated time by Alamar Blue reduction. Cell growth values at each time point were normalized to baseline values of untreated cells at Day 0. The effects of treatments on cell growth in each line were first analyzed in Excel using ANOVA. Post-hoc comparisons were performed between specific treatments using two-sided t tests if the overall difference reached statistical significance. Significance was set at a p value of 0.05 for all comparisons.

FISH Studies

Human - specific pan centromeric DNA was labeled with Cy3, mouse - specific pan centromeric DNA was labeled with either Cy3 or FITC (All from Cambio Ltd, UK). The experiment was done using vender's kit and protocol. Briefly, five-micrometer paraffin-embedded tumor tissue sections were deparaffinized in xylene, digested in 80μg/ml of pepsin at 37°C for two hours, and then washed in PBS, the tissue sections were denatured in 70% of formamide/2xSSC at 80°C for 3 min. The probes were denatured at 80°C for 10 min and were then applied to the pretreated tissue sections, coverslipped, and sealed for hybridization at 37°C over night. The unbound probes were washed off in 50% formamide/2XSSC at 37°C. The sections were counterstained with DAPI.

CK5 Immunohistochemistry

An antibody for Cytokeratin 5 (clone EPR1600Y, Abcam) was applied at a 1:200 dilution overnight after antigen retrieval in Tris/ EDTA buffer pH 9.0 @ 95C in a steamer. Secondary antibody, species specific, was performed using an HRP labeled polymer which is conjugated with secondary antibodies (Envision–HRP, Dako). Staining was completed with 3′3'diaminobenzidine as the chromagen. Counter staining accomplished using Mayer's Hematoxylin (Biogenex).

SUPPLEMENTAL REFERENCES

Feng, Y., Nie, L., Thakur, M.D., Su, Q., Chi, Z., Zhao, Y., and Longmore, G.D. (2010). A multifunctional lentiviral-based gene knockdown with concurrent rescue that controls for off-target effects of RNAi. Genomics Proteomics Bioinformatics *8*, 238–245.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics *7*, 96.

Rozen S. and Skaletsky H., eds. (2000). Primer3 on the WWW for general users and for biologist programmers (Totowa, NJ: Humana Press).

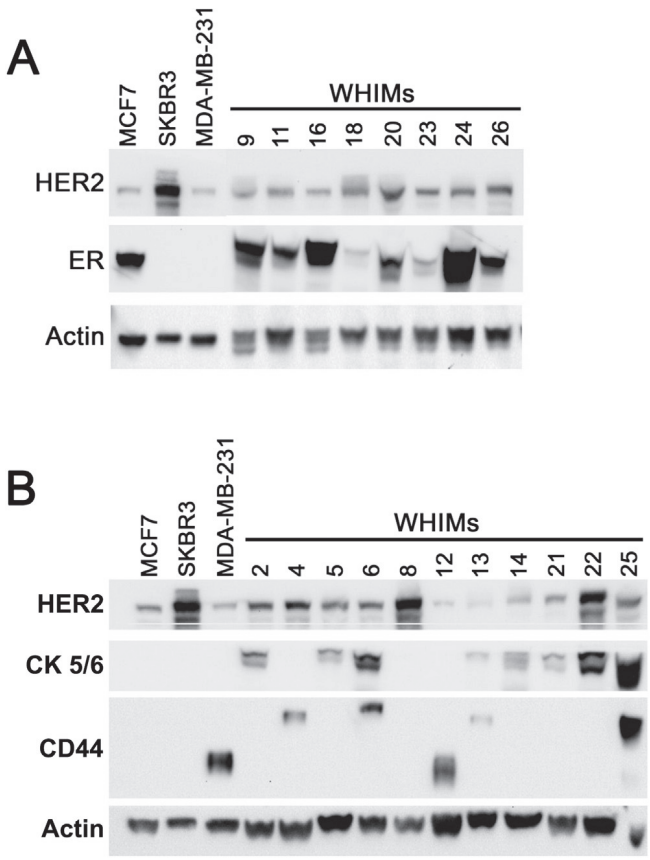

Figure S1. Protein Biomarker Expression across the Breast Cancer PDX Panel, Related to Results

(A and B) Tumor lysates from ER-positive (A) and ER-negative (B) WHIM lines were analyzed by western blot to confirm ER and HER2 status. Lysates from ER-negative WHIM lines were also probed with antibodies against the cytokeratin 5/6 (CK 5/6) and CD44 markers. The MCF7 (ER-positive), SKBR3 (ER-negative, HER2-positive) and MDA-MB-231 (ER-negative, CD44-positive) breast cancer cell lines were included as blotting controls.

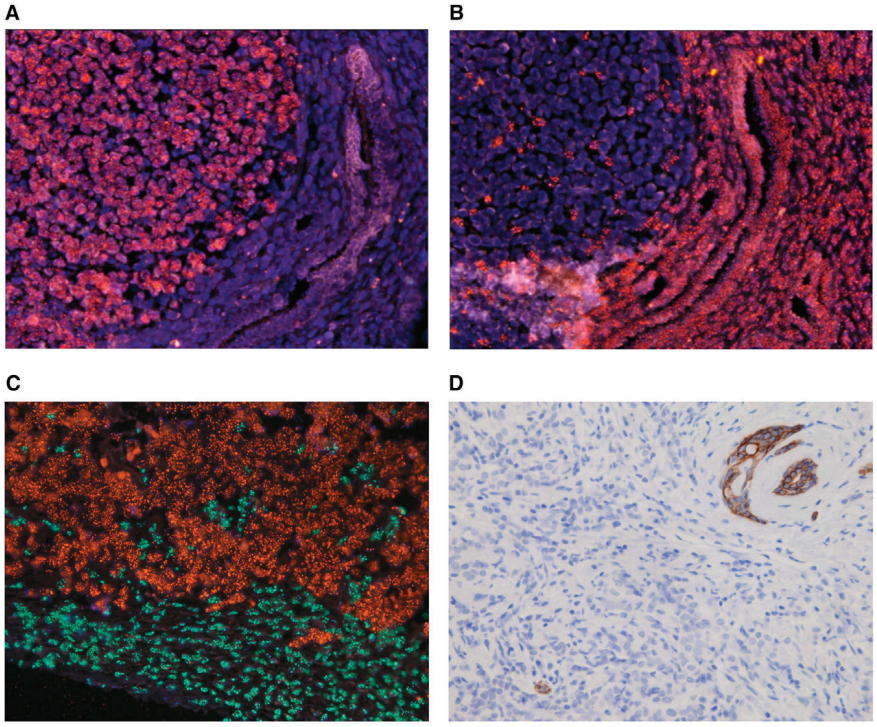

Figure S2. In Situ Studies of WHIM Lines and Progenitor Samples, Related to Results

(A) WHIM2 ovary met hybridized with human specific centromere probe.

(B) WHIM2 ovary met hybridized with mouse-specific centromere probes on adjacent section.

(C) WHIM6 hybridized with both human (orange color) and mouse (green color) specific centromere probes showing absence of cells that have both human and mouse hybridization signals indicating the absence of inter-species cell fusion events.

(D) Human-specific CK5 staining of a lumenal progenitor sample for WHIM 20 showing normal epithelial cells positive for CK5 "trapped" among malignant cells that are CK5 negative, thus explaining why CK5 mRNA is found in human progenitor samples but not in the xenograft as the normal epithelial cells do not grow in the PDX.

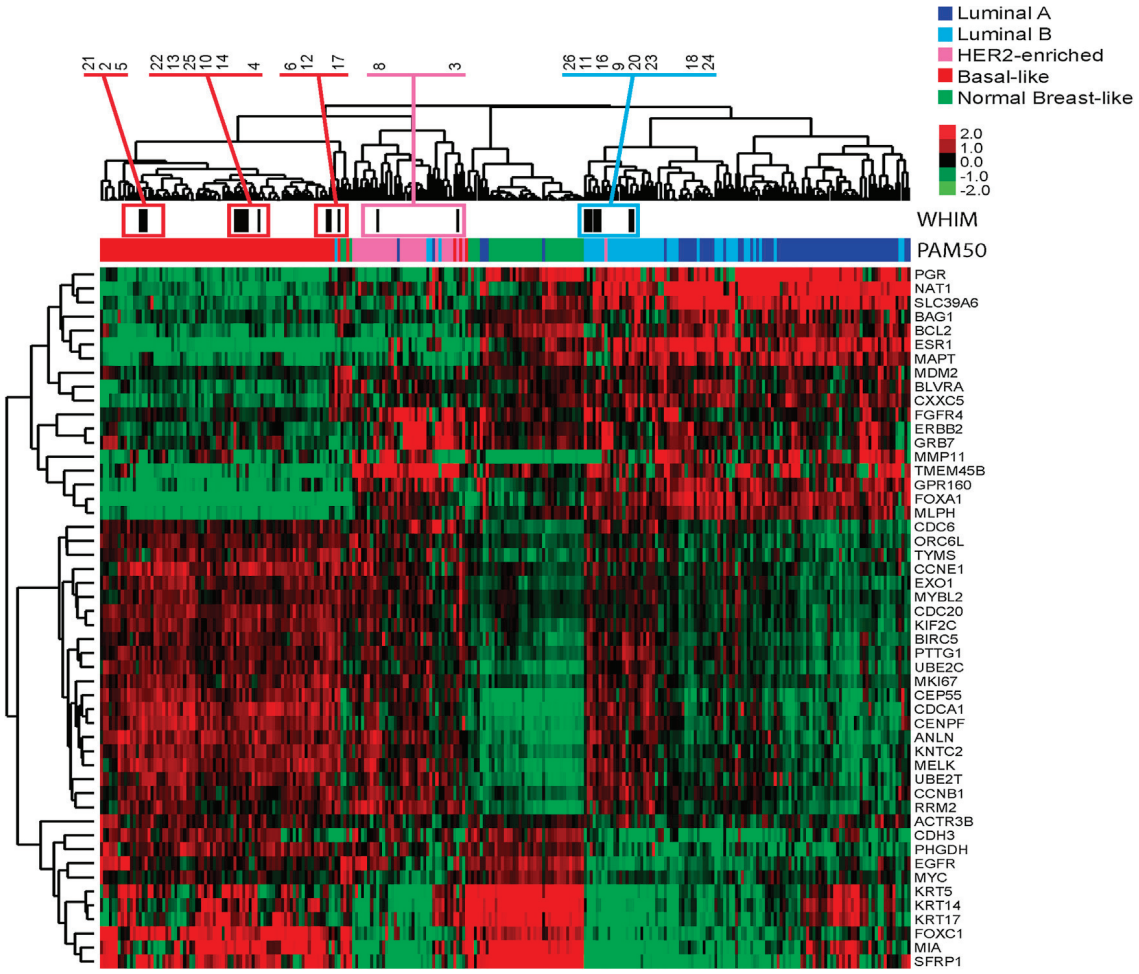

Figure S3. Breast Cancer Intrinsic Molecular Subtyping of the WHIM Models, Related to Results

Coclustering of 250 human breast samples representing all the PAM50 intrinsic subtypes and 22 HIM models using the 50 PAM50 genes. Gene expression data of all samples has been obtained using 244K Agilent microarrays.

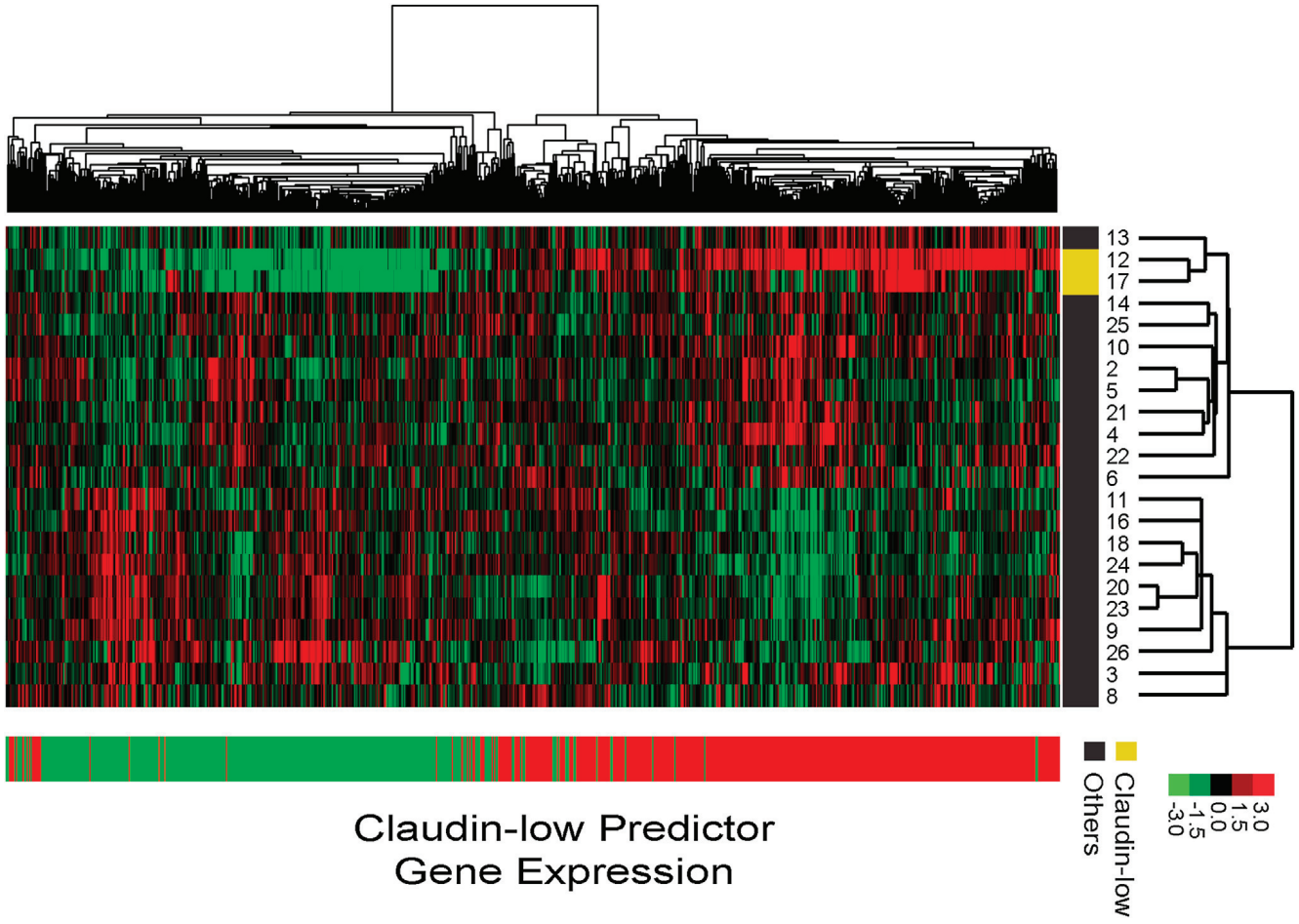

Figure S4. Expression of Our Previously Reported Claudin-Low Signature of 807 Genes across the 22 WHIM Models, Related to Results and Figure 1

On the right, the expression of each gene (up- or downregulated) of the Claudin-low signature is shown. The two Claudin-low samples (WHIM 12 and WHIM 17) are identified below the array tree.

Figure S5. RPPA Data across the Panel of Breast Cancer PDX, Related to Results and Table 1

The RPPA data for WHIM samples and 386 TCGA samples were combined after standardizing the data for each marker (i.e., subtracting the mean and then divided by the standard deviation) in the separate data sets. The combined samples were hierarchically clustered using a Pearson correlation based distance matrix ((1-rho)/2, where rho is the Pearson correlation matrix) and the "ward" linkage based on Ward's minimum variance. In every case the samples from each WHIM line clustered adjacently, including the two double PDX model isolations (WHIM 2 and 5, WHIM 20 and WHIM 23). This demonstrates that intra-PDX heterogeneity was considerably less than the inter-tumoral heterogeneity in a large RPPA data set and was relatively stable over time and passage.

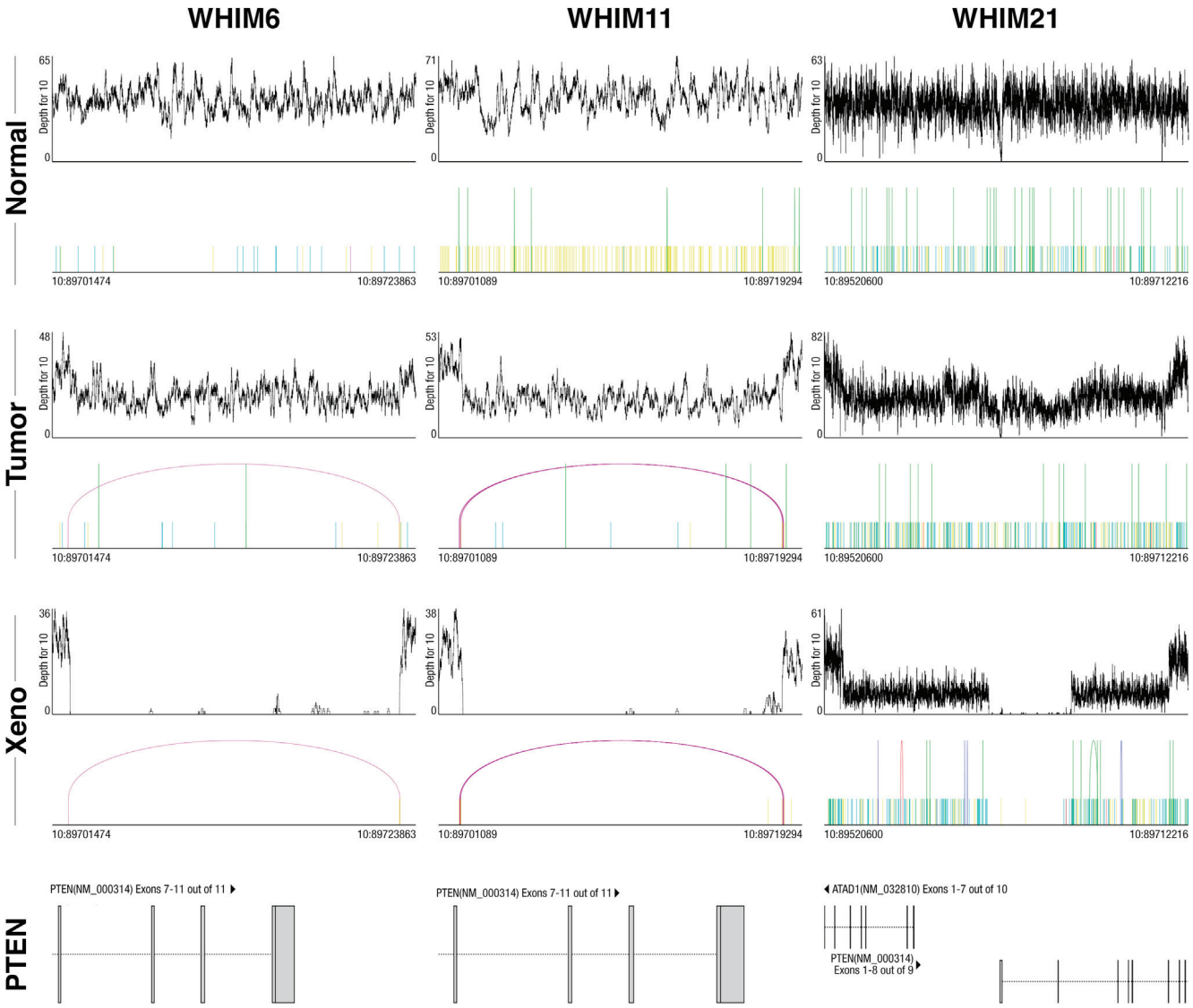

Figure S6. Computational Purification of Human Tumor DNA Sequence by Informatic Removal of Murine Sequence Arising from the Xenograft Stroma, Related to Results and Figure 2

Detection of biallelic PTEN deletions is enhanced in the PDX SV analysis due to removal of murine sequence before somatic mutation and SV analysis (computational purification of human tumor sequence). In each of the 3 examples illustrated (WHIM6. WHIM11 and WHIM21), the loss of read counts due to biallelic deletions are revealed with clarity. In contrast, contamination of the human originating tumor with normal DNA prevents a conclusion regarding biallelic loss, particularly when break-spanning reads were not detected (WHIM 20).

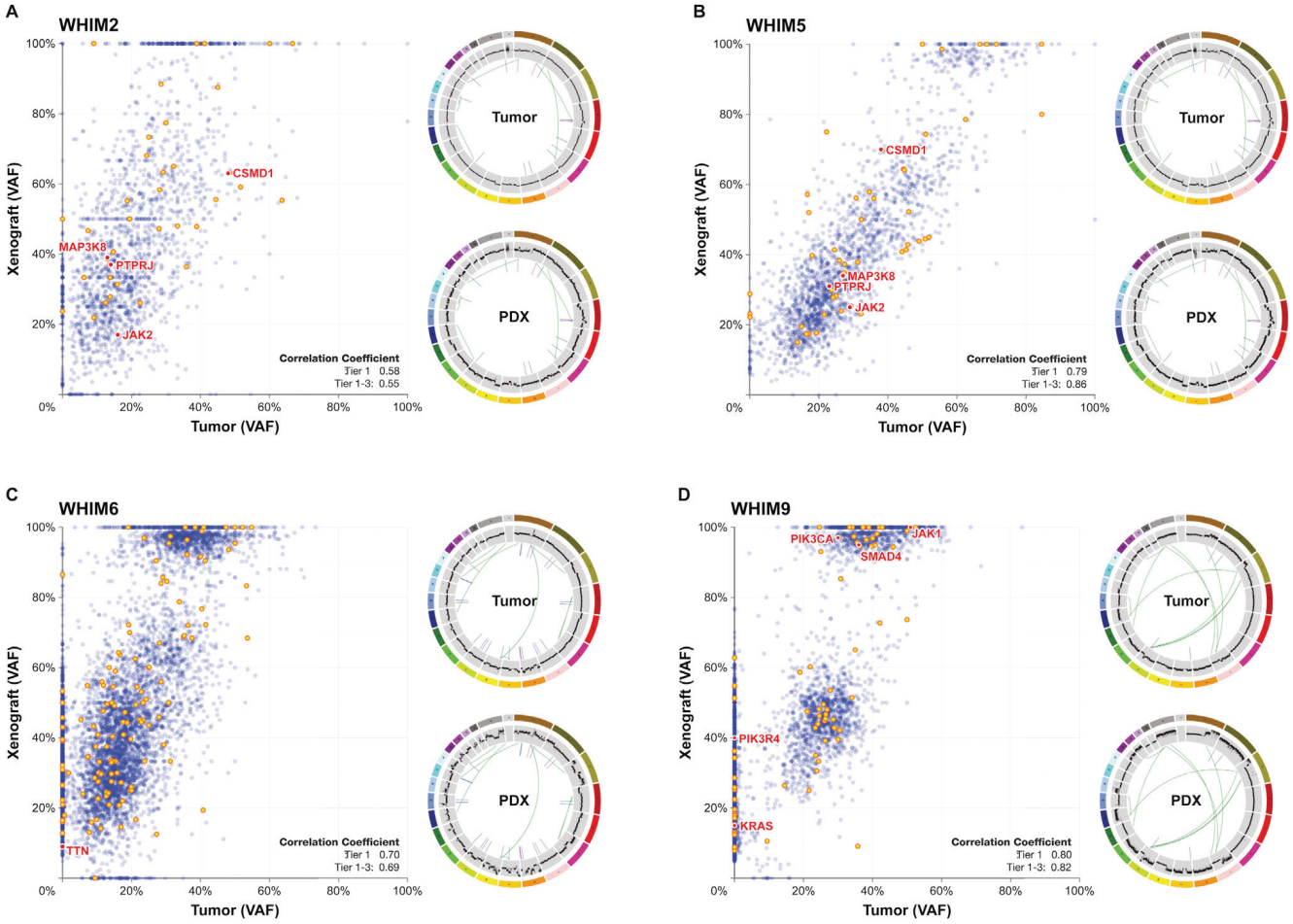

Figure S7. The Circos Plots and Variant Allele Frequency Plots for Four WHIM PDX Models Subjected to Whole-Genome Sequencing, Related to Results

(A) WHIM2, (B) WHIM5, (C) WHIM6, (D) WHIM9. Overall the Circos plots show closely matched SV and CNV in the tumor of origin and the paired WHIM line. To compare differences in mutant allele frequency between the originating tumor and the PDX counterpart, the read counts for each mutant and wild-type allele were expressed as a percentage of all reads at that position and analyzed by scatter plot and simple correlation coefficient. These show considerably more variation in the human to PDX comparisons but variant allele frequencies are, nonetheless, often preserved. High-resolution images of Figures S7–S9 can be viewed at http://digitalcommons.wustl.edu/hamlet_spotlight/.

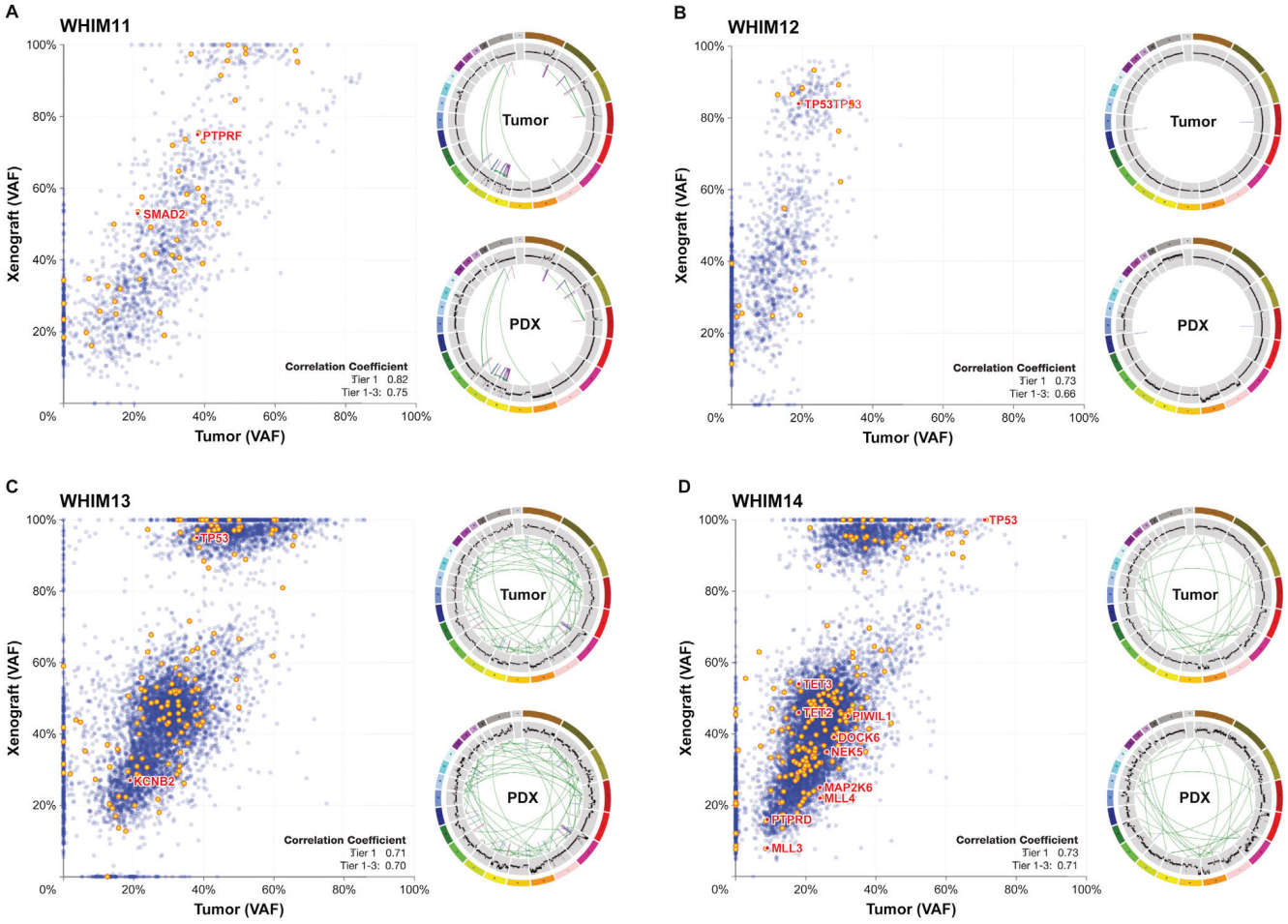

Figure S8. The Circos Plots and Variant Allele Frequency Plots for Four WHIM PDX Models Subjected to Whole-Genome Sequencing, Related to Results

(A) WHIM11, (B) WHIM12, (C) WHIM13, (D) WHIM14. Overall the Circos plots show closely matched SV and CNV in the tumor of origin and the paired WHIM line. To compare di_erences in mutant allele frequen-cy between the originating tumor and the PDX counterpart, the read counts for each mutant and wild-type allele were expressed as a percentage of all reads at that position and analyzed by scatter plot and simple correlation coe_cient. These show considerably more variation in the human to PDX comparisons but variant allele frequencies are, nonetheless, often preserved. High-resolution images of Figures S7–S9 can be viewed at http://digitalcommons.wustl.edu/hamlet_spotlight/.

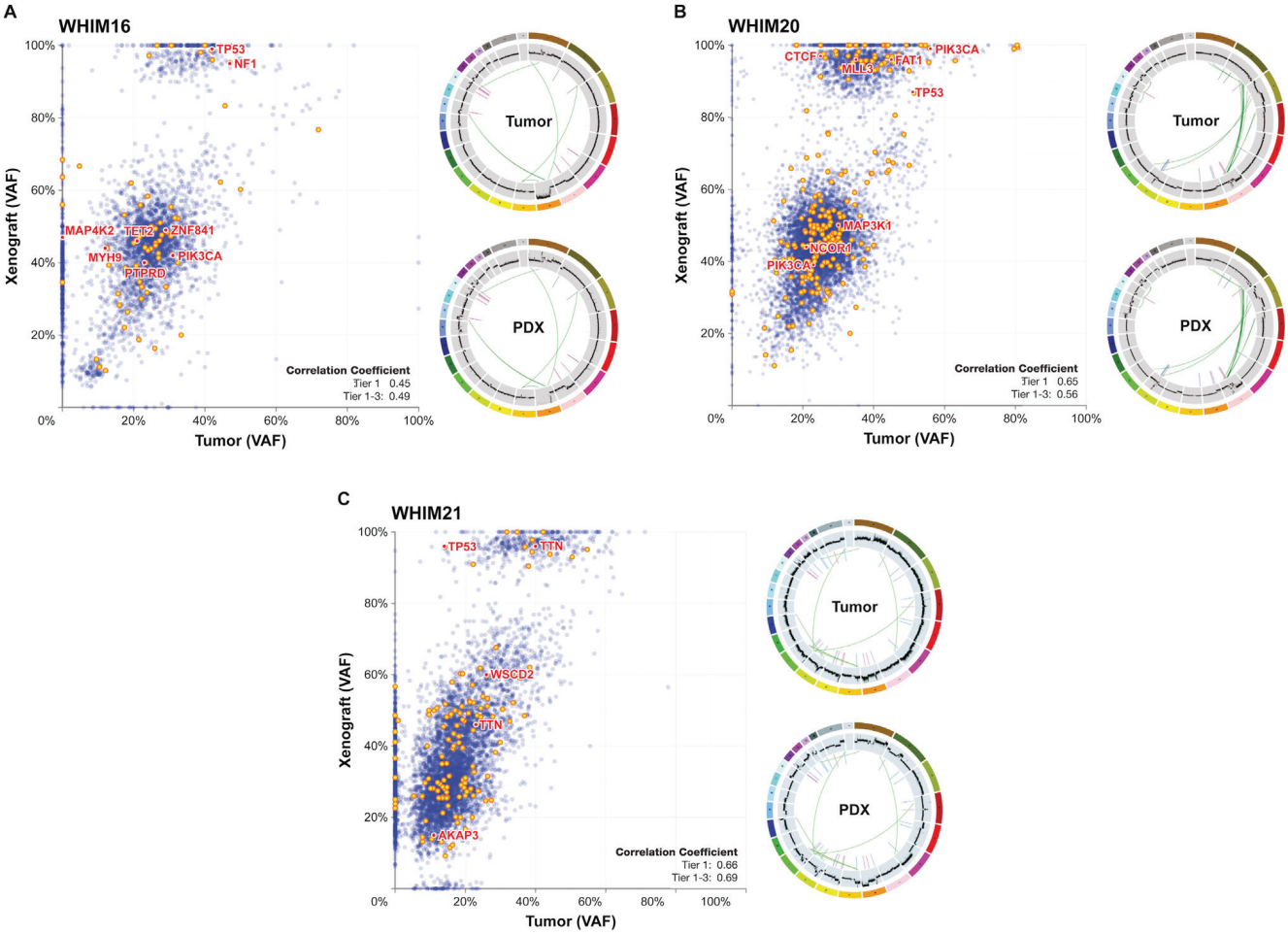

Figure S9. The Circos Plots and Variant Allele Frequency plots for Three WHIM PDX Models Subjected to Whole-Genome Sequencing, Related to Results

(A) WHIM16, (B) WHIM20, (C) WHIM21. Overall the Circos plots show closely matched SV and CNV in the tumor of origin and the paired WHIM line. To compare differences in mutant allele frequency between the originating tumor and the PDX counterpart, the read counts for each mutant and wild-type allele were expressed as a percentage of all reads at that position and analyzed by scatter plot and simple correlation coefficient. These show considerably more variation in the human to PDX comparisons but variant allele frequencies are, nonetheless, often preserved. High-resolution images of Figures S7–S9 can be viewed at http://digitalcommons.wustl.edu/hamlet_spotlight/.

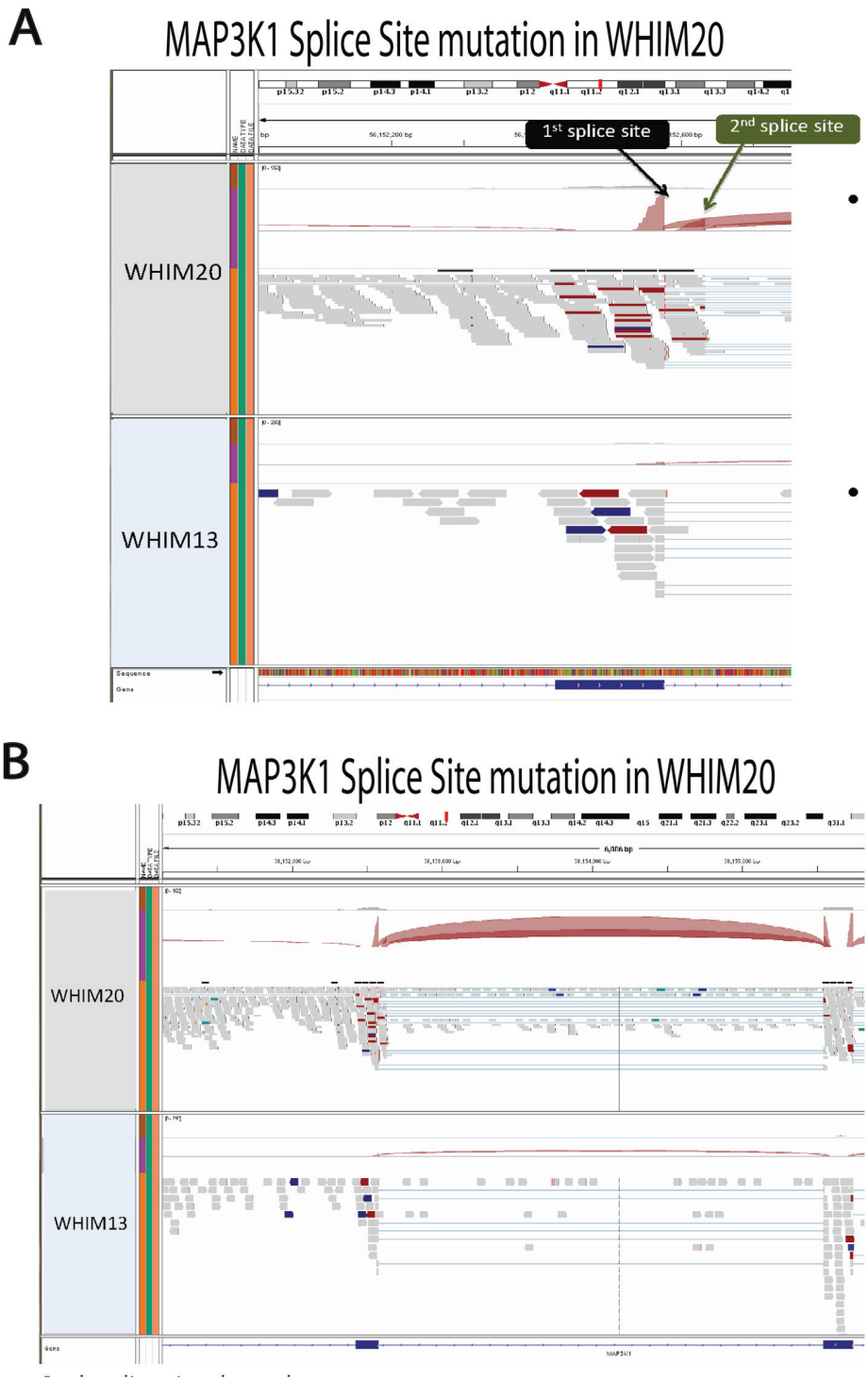

Figure S10. Alternative MAP3K1 Splicing in WHIM20, Related to Results

(A) IGV screenshot of RNA-Seq coverage for MAP3K1 (blue bar) in WHIM20 (top) and WHIM13 (bottom). Both WHIM20 and WHIM13 show the canonical splice site, referred to as '1st splice site', whereas a second splice site within the intron is utilized only in WHIM20, referred to as '2nd splice site'. Unlike the canonical splice site, the splice site alters the open reading frame.

(B) An IGV screenshot showing the a larger genomic region of MAP3K1 demonstrates that both splice variants utilize the same acceptor site.

A

ESR1-YAP1 **6 unique fusions** **WHIM18** **Inter-chromosomal**

5' ESR1	1-365	Includes DNA binding domain, but not steroid-binding domain
3' YAP1	231-504	Includes transactivation domain

FANK1-JMJD1C **6 unique fusions** **WHIM11** **Intra-chromosomal**

5' FANK1	1-4	Excludes all UniProtKB annotated domains
3' JMJD1C	1-2,358	Includes entire protein. Has JmjC-domain (JmjC-domain proteins may be protein hydroxylases that catalyse a novel histone modification)

ACSBG2-KIAA0319L **3 unique fusions** **WHIM5** **Inter-chromosomal**

5' ACSBG2	1-99	Excludes all UniProtKB annotated domains .
3' KIAA0319L	223 - 1,049	Excludes cytoplasmic and extracellular domain. Includes all five PKD (Polycystic Kidney Disease) domains

C1orf182-MLL3 **3 unique fusions** **WHIM18** **Inter-chromosomal**

5' C1orf182		0 amino acids (fusion only includes untranslated exons)
3' MLL3		845-4,911

ADARB2-DIP2C **1 unique fusions** **WHIM13** **Intra-chromosomal**

5' ADARB2	1-33	Does not include any annotated domains
3' DIP2C	29-1,556	no UniProtKB-annotated domains

B

ADAM18-ADAM2 **6 unique fusions** **WHIM14** **Converging Genes**

5' ADAM18	1-34	Includes signal peptide.
3' ADAM2	191-735	Excludes signal and propeptide domains

SNRPC-ANKS1A **3 unique fusions** **WHIM13** **Read-through**

5' SNRPC	1-53	Includes zinc finger
3' ANKS1A	21-570	no UniProtKB-annotated domains

CLTC-TMEM49 **2 unique fusions** **WHIM13** **Read-through**

5' CLTC	1-1,609	Includes all UniProtKB-annotated domains
3' TMEM49	169-309	no UniProtKB-annotated domains

Figure S11. In-Frame Gene Fusions, Related to Results

Gene fusions found to be in-frame are sorted in descending order of their frequency. Predicted protein domain structures shown using the amino acid coordinates for the largest predicted fusion product.

(A and B) Fusions are dividing into (A) genomic events (interchromosomal and intrachromosomal events) and (B) putative transcription events (converging genes and read-through transcription between adjacent genes in the same orientation).

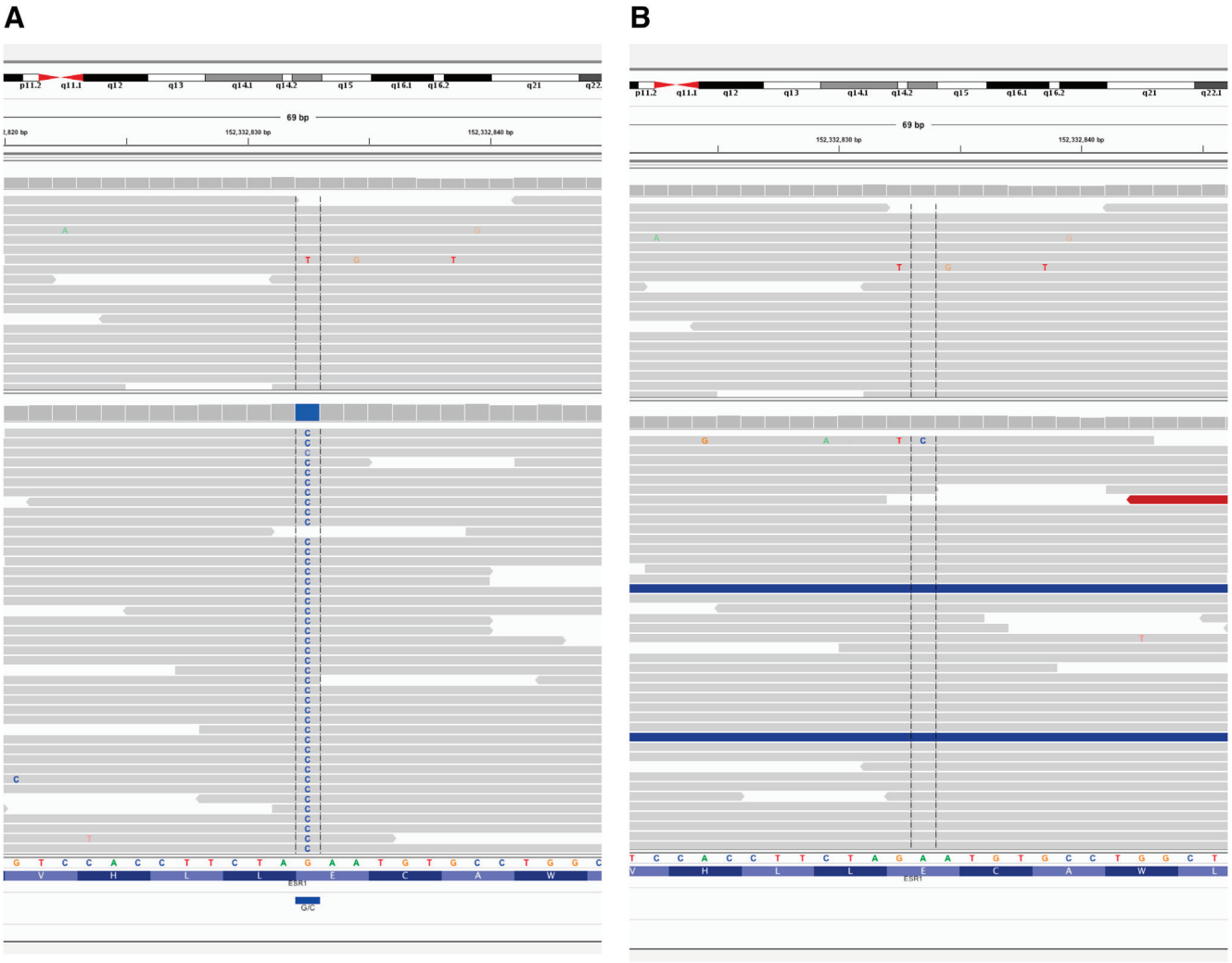

Figure S12. Integrated Genomics Viewer Showing the Mutation that Generated the E380Q Mutation in WHIM24, Related to Results and Table 1

(A and B) While 100% of the reads in WHIM24 are the mutant allele (A), only one read is mutant in the originating tumor (B), indicating strong enrichment during the engraftment process.

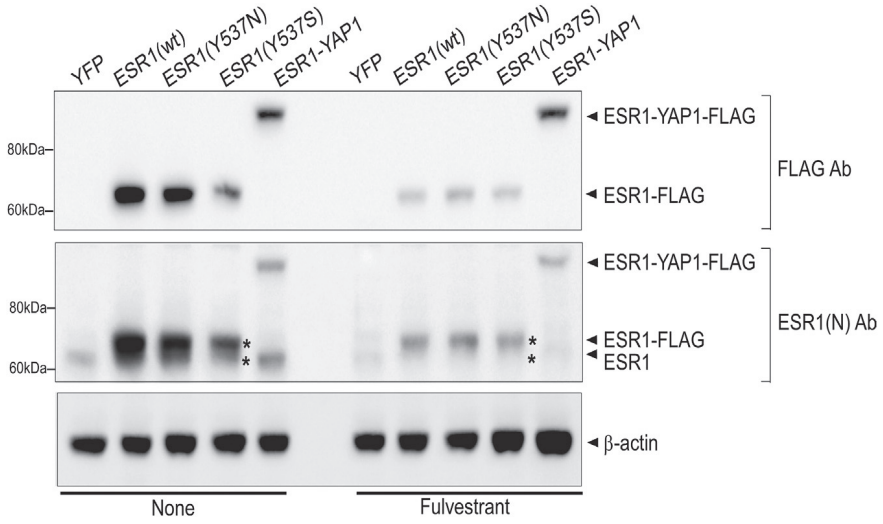

Figure S13. ER-YAP1 Fusion Protein Is Resistant to Degradation Induced by Fulvestrant, Related to Results and Figure 6

T47D cells stably expressing YFP, wild-type ER, ER point mutants (Y537N and Y537S) and ER-YAP1 fusion were cultured in charcoal-stripped serum for 8 days in the absence or presence of 0.5 μM fulvestrant and analyzed by western blot using antibodies against the FLAG-tag (fused to all four exogenously expressed ER proteins) or N-terminus of ER. Actin was blotted as the loading control.

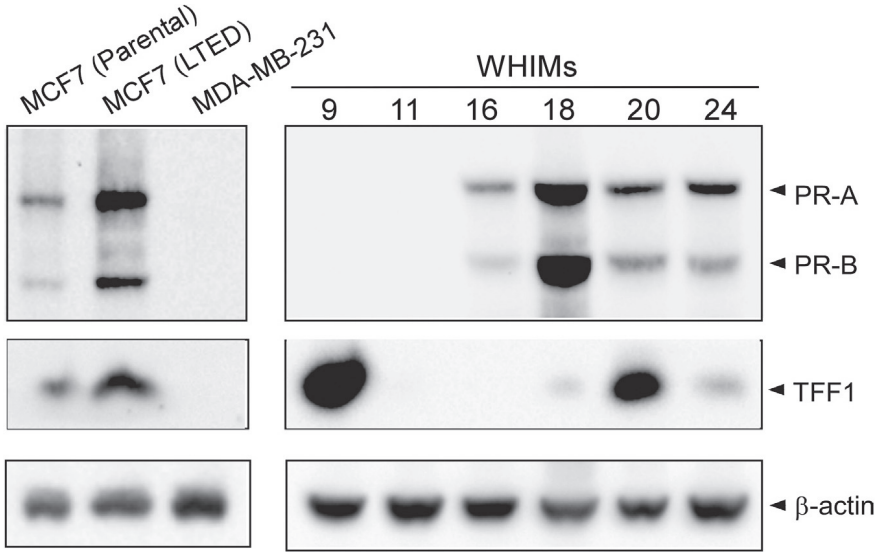

Figure S14. Expression of Progesterone Receptor and TFF1 in ER+ WHIM Tumors, Related to Results and Figure 2

Six ER+ WHIM tumors were analyzed for PR and TFF1 expression levels by western blot. Three breast cancer cell lines were probed in parallel as positive (MCF7 and MCF7-LTED) and negative (MDA-MB-231) controls. Actin was used as the loading control.

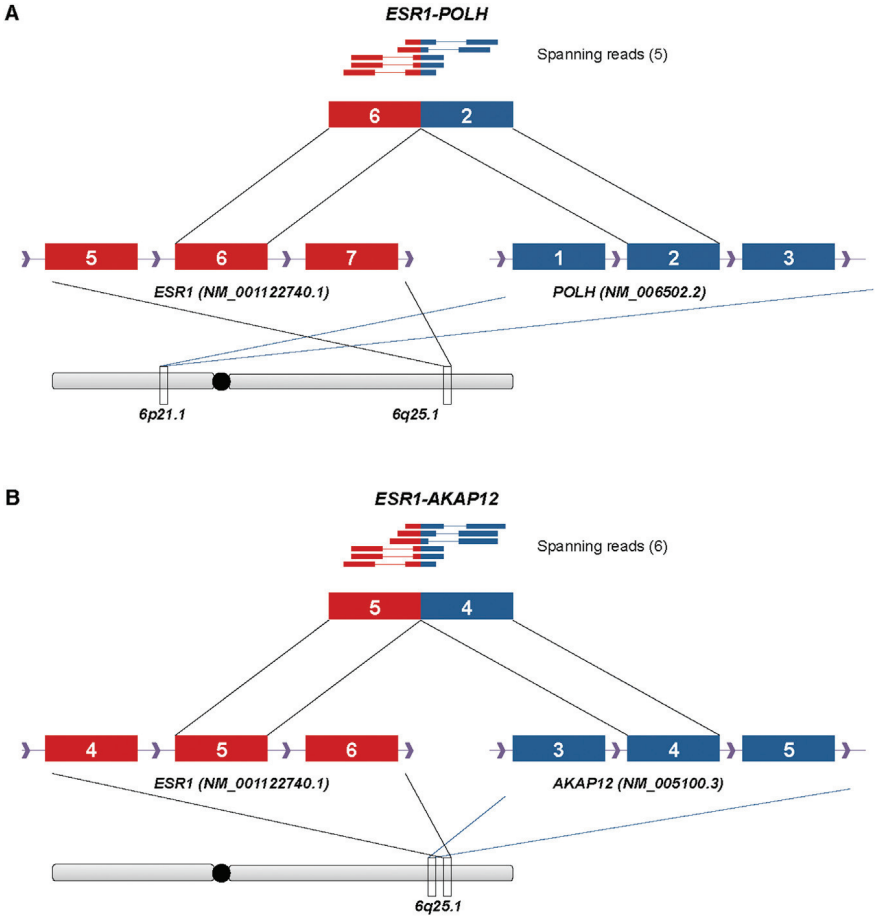

Figure S15. ESR1 Fusion Genes Observed in the TCGA RNA-Seq Data Set, Related to Discussion and Figure 2

(A and B) Additional examples of ESR1 fusion genes to complement the ESR1-YAP1fusion in WHIM18. Analysis of the Cancer Genome Atlas Network mRNAseq data revealed two in-frame ESR1 fusion genes that preserved the N-terminal DNA binding and AF1 domains of ESR1 but replacing the C-terminal binding and AF2 domains of ESR1 with sequence from AKAP12 or POLH.