

**SUPPLEMENTARY INFORMATION**  
**(ONLINE SUPPORTING INFORMATION)**

**Characterization of the genomic features and expressed fusion genes in  
micropapillary carcinomas of the breast**

Natrajan et al.

Supplementary Methods

Supplementary Figures S1-S6

Supplementary Tables S1-S7

## SUPPLEMENTARY METHODS

### Tumor samples

Two cohorts of micropapillary carcinomas (MPCs) were analyzed; the first cohort comprised 16 consecutive formalin fixed paraffin embedded (FFPE) MPCs, 11 pure and 5 mixed, which were retrieved from the authors' institutions (Table 1), and a second, validation cohort comprised 14 additional consecutive FFPE MPCs, retrieved from Molinette Hospital, Turin, Italy. Frozen samples were available from five out of the 16 cases from the first cohort of MPCs.

As a comparator for the results of the Sequenom mutation profiling, a cohort of 16 consecutive IC-NSTs matched to the first cohort of 16 MPCs according to ER and HER2 status and histological grade were retrieved from a series of breast cancers previously analyzed by aCGH[1]. In addition, 14 IC-NSTs matched according to grade, and ER and HER2 status to tumors from the second cohort of 14 MPCs, and 48 grade 3 IC-NSTs were retrieved from Hospital La Paz, Madrid, Spain[1] (Supplementary Table S1).

### Power calculation

For power calculations, we have assumed that if MPCs were driven by a recurrent fusion gene in a way akin to secretory carcinomas (which harbor the *ETV6-NTRK3* fusion gene in >95% of cases[2-4]) or adenoid cystic carcinomas of the breast (which harbor the *MYB-NFIB* fusion gene in >90% of cases[5]), a 'pathognomonic' driver event would be present in at least  $\geq 70\%$  of cases (an estimate that is conservative). Therefore, based on a binomial distribution, by sequencing 5 samples we would have been able to identify a recurrent event (i.e. in two or more cases) with 97% statistical power. Furthermore, a 'pathognomonic' driver event present in 50% of cases would be detectable with 80% power.

### Immunohistochemical analysis

Representative sections of each case were cut at 3 $\mu$ m and mounted on silane-coated slides. Immunohistochemistry was performed as previously described [6,7], using antibodies raised against estrogen receptor (ER), progesterone receptor (PR), HER2, and epithelial membrane antigen (EMA). Antibody clones, dilution, antigen retrieval methods, scoring systems, and cut-offs [8,9] are summarized in Supplementary Table 2.

Positive and negative controls (omission of the primary antibody and IgG-matched serum) were included for each immunohistochemical run. The scoring was performed by at least two pathologists (CM, AS and/ or JSR-F).

#### **Microdissection, DNA extraction, RNA extraction**

Representative 8 $\mu$ m-thick sections of the MPCs and IC-NSTs were subjected to microdissection with a sterile needle under a stereomicroscope (Olympus SZ61, Tokyo, Japan) to ensure a percentage of tumor cells greater than 90%, as previously described[7,10]. DNA was extracted from the 16 MPCs and 16 grade-, ER-, and HER2-matched IC-NSTs using the DNeasy Blood and Tissue Kit (Qiagen, Crawley, UK). In addition, for 8 MPCs matched adjacent normal breast tissue was microdissected and subjected to DNA extraction. Double-stranded DNA concentration and DNA quality were determined using a Qubit Fluorometer (Invitrogen, Paisley, UK) and a multiplex PCR, respectively, as previously described[6,10]. RNA was extracted from the microdissected MPCs and IC-NSTs FFPE samples using the RNeasy kit (Qiagen). Frozen samples from five MPCs and 48 grade 3 IC-NSTs were cut, microdissected, and subjected to RNA extraction with Trizol (Invitrogen), according to the manufacturer's protocol. RNA quantity and quality was assessed using Agilent 2100 Bioanalyzer with RNA Nano LabChip Kits (Agilent Biosystems, Stockport, UK).

#### **Microarray comparative genomic hybridization (aCGH)**

aCGH data were pre-processed and analyzed using the Base.R script in R version 2.14.0, as previously described[1,11]. Genomic DNA from each sample was hybridized against a pool of normal female DNA derived from peripheral blood. Raw Log<sub>2</sub> ratios of intensity between samples and pooled female genomic DNA were read without background subtraction and normalized in the LIMMA package in R using PrinTipLoess. Outliers were removed based upon their deviation from neighboring genomic probes, using an estimation of the genome-wide median absolute deviation of all probes. Log<sub>2</sub> ratios were rescaled using the genome wide median absolute deviation in each sample and then smoothed using circular binary segmentation (cbs) in the DNACopy package [12-15]. After filtering polymorphic BACs and BACs mapping to chromosome Y, a final dataset of 31,157 clones with unambiguous mapping information according to build hg19 of the human genome (<http://www.ensembl.org>). A categorical analysis was applied to the BACs after classifying them as representing amplification (>0.45), gain (>0.08 and

$\leq 0.45$ ), loss ( $< -0.08$ ) or no change, according to their cbs-smoothed log<sub>2</sub> ratio values [7,14]. Threshold values were determined and validated as previously described [13,14].

### **Mutation screening and validation**

Sixteen MPCs and 16 grade-, ER- and HER2-matched IC-NSTs were subjected to mutation screening using the OncoCarta Panel v 1.0 (Sequenom, San Diego, CA, USA), detecting 238 mutations in 19 common cancer-related genes, as previously described[12,15]. The prevalence of mutant alleles was estimated by calculating the ratio of the area of the raw spectra of the mutant allele to its wild-type allele. Mutations were validated using Sanger sequencing as previously described[12,15]. Primer sequences are available at <http://rock.icr.ac.uk/collaborations/Mackay/Micropapillary>. Sequences were visualized using 4Peaks (<http://4peaks.en.softonic.com/>).

Eight MPCs, from which both primary tumor and normal breast tissue could be microdissected, were subjected to microdissection, and DNA was extracted as previously described[7,10]. Tumor and germline DNA were subjected to targeted capture massively parallel sequencing using a platform containing baits targeting all exons of 273 genes that were either recurrently mutated in breast cancer or are involved in DNA repair pathways (Supplementary Table S3). Custom oligonucleotides (NimblegenSeqCap) were designed for hybridization capture of all protein-coding exons of 273 genes (Supplementary Table S3). Barcoded sequence libraries were prepared (New England Biolabs, KapaBiosystems) using 50ng DNA and pooled at equimolar concentrations into a single exon capture reaction as previously described[16,17]. Sequencing was performed in a single lane of an Illumina HiSeq2000 (San Diego, CA), and reads were aligned to the reference human genome hg19 using the Burrows-Wheeler Aligner (BWA) [6,18]. Somatic single nucleotide variants were identified using a combination of muTect[19], MutationSeq[20], Haplotypecaller[21] and VarScan2[22]; only single nucleotide variations identified by at least two out of the four callers employed were considered as valid. This conservative approach has been employed to minimize false positive results obtained with high depth targeted massively parallel sequencing performed with DNA extracted from formalin-fixed, paraffin-embedded tissues. For small insertions and deletions (indels), Haplotype Caller and VarScan2 were employed. All candidate mutations were reviewed manually using the Integrative Genomics Viewer[23]. Mutations with allelic frequency of  $< 1\%$  and/ or supported by  $< 5$

reads were disregarded. Regions with loss of heterozygosity were identified using exomeCNV[24].

### **Paired-end massively parallel RNA sequencing**

Briefly, messenger RNA was selected using oligo-dT magnetic beads from 3µg of total tumor RNA. RNA was fragmented at 94°C for 5 minutes in fragmentation buffer (Illumina) and converted to single stranded cDNA using SuperScript II reverse transcriptase, followed by second-strand cDNA synthesis using Escherichia coli DNA polymerase I. Double-stranded cDNA was end repaired by using T4 DNA polymerase and T4 polynucleotide kinase; monoadenylated using a Klenow DNA polymerase I (3'-5' exonucleotide activity), and adapters ligated using T4 DNA ligase. The adaptor-ligated cDNA library was then fractionated on a 2% agarose gel, and a smear corresponding to 200 nucleotides was excised, purified, and PCR amplified using high fidelity Phusion DNA polymerase (Finnzymes, Vantaa, Finland). The library was quantified using the Agilent 1000 kit on the Agilent 2100 Bioanalyzer following the manufacturer's instructions. 3pM was then loaded onto each lane of a flowcell, and sequencing performed with and 2x 54bp cycles on the Genome Analyser II.

### **Reverse-transcription PCR (RT-PCR), PCR and Sanger sequencing validation**

Breakpoints of validated in-frame fusion genes were verified at the genomic level using the Expand Long Template PCR system (Roche Diagnostics Ltd, Burgess Hill, UK). For each fusion gene pair, the maximum size of intervening intronic sequence was calculated and primers were designed for the amplicon to encompass the fusion junction. Reverse transcription was performed with Superscript III (Invitrogen) using 500ng of RNA per reaction, according to the manufacturer's instructions. RT-PCR was performed using AmpliTaq Gold (Applied Biosystems, Warrington, UK) as previously described [25]. Products were sequenced using a DNA Sequencing Kit BigDye Terminator v 1.1 Cycle Sequencing Ready Reaction Mix (Applied Biosystems), as previously described[13]. Sequences were visualized using 4Peaks (<http://4peaks.en.softonic.com/>).

Validated in-frame fusion genes (i.e. *SLC2A1-FAF1* and *BCAS4-AURKA*) and out-of-frame fusion genes with recurrent partners found in independent datasets [26-29](i.e. *ELMO2-RAE1* and *LASP1-CDK12*) were screened at the cDNA level in an independent

cohort of MPCs (n=14), grade- and ER-matched IC-NSTs (n=14) and grade 3 IC-NSTs (n=48) using RT-PCR. Quantitative real-time PCR (qRT-PCR) assays for *BCAS4-AURKA* and *SLC2A1-FAF1* were performed as previously described[27], using SYBR green master mix (Applied Biosystems) and run on the 7900 Real time PCR device (Applied Biosystems). Relative mRNA levels were normalized to the expression of  $\beta$ -actin as previously described[27].

### **Cell line models**

The ER-positive breast cancer cell lines, MCF7, BT474, T47D and ZR75.1 were obtained from ATCC (LGC Standards, Teddington, UK). The *HER2*-amplified cell lines JIMT1, UACC812, UACC893, VP229, SKBR3, ZR75.30, HCC1569, HCC1954, MDA-MB-453, MDA-MB-361 were obtained from ATCC and the cell lines SUM225, SUM190 were obtained from Asterand, Herts, UK). All cell lines were and maintained as previously described[30,31]. Cell lines were authenticated using short tandem repeat (STR) genotyping using the PowerPlex 1.2 STR system (Promega UK, Southampton, UK), according to the manufacturer's instructions.

### **Functional assessment of in-frame fusion genes**

The *SLC2A1-FAF1* and *BCAS4-AURKA* fusion open reading frames (ORFs) were PCR amplified from primary tumor MPC10 and cloned into a mammalian expression vector pCMVentry, with a C-terminal DDK tag (Origene). Cloned fusions, full-length 3-prime partner gene constructs or empty vectors were transfected into in four ER-positive breast cancer cell lines (MCF7, BT474, T47D and ZR75.1) using Lipofectamine 2000 (Invitrogen). Antibiotic selection using geneticin (Invitrogen) was performed 48 hours after transfection for 5-7 days, before cells were then seeded in 96 well plates for growth assays. Cell population was estimated every 24 hours, for 9 days, using the CellTiter-Glo® cell viability assay (Promega), as previously described[32], with each reading normalized to day 1 to determine the fold change. Western blotting was performed to confirm expression of DDK in transfected cell lines prior to performing growth assays as previously described[32].

### **Analysis of RAD51 foci formation**

In brief, cells were grown on coverslips (12mm, Thermo Scientific) and exposed to 10 Gray (Gy) of ionizing radiation. After 4-6 hours recovery, cells were fixed, permeabilized,

and immunostained with primary antibodies targeting RAD51 (polyclonal rabbit antibody, Santa Cruz) and detected with alexa fluor 488 (Invitrogen). Nuclei were counterstained with 4',6-diamidino-2- phenylindole (DAPI). The presence of RAD51 foci was evaluated in a minimum of 100 cells in three independent experiments with an automated inverted fluorescence microscope (LEICA – DMI6000 B). RAD51 foci formation was scored as a percentage of cells with phospho- $\gamma$ -H2AX foci formation without treatment, 6 hours after 10 Gy of irradiation, and with 1nM BMN673 for 24 hours. A minimum of 100 cells was counted to determine RAD51 focus formation frequency in three independent experiments.

### **Long-term survival assay**

Long-term survival assays were performed as previously described[32,33]. In brief, cells were seeded in six-well plates in triplicates at a concentration of 1000-2000 per well and treated with the inhibitor after 24 hours, being continuously exposed to the drug ( $10^{-11}$  to  $10^{-7}$  M) dissolved in dimethyl sulfoxide (DMSO) (control: DMSO). Media and drug were replaced every 3 days. After 14 days, cells were fixed with 10% tri-chloro-acetic acid (TCA) and stained with sulforhodamine B (SRB) (Sigma) and a colorimetric assay was performed as described previously[32,33]. The PARP inhibitor BMN673 was a kind gift from BioMarin Pharmaceuticals, USA, and olaparib (AZD2281/KU0058948) was obtained from SelleckBio (Munich, Germany).

### **Statistical analyses**

Growth curves and drug sensitivity curves were constructed using non-linear regression (curve fit) analyses in Prism v5.04 (Graphpad Software Inc, La Jolla, CA, USA). Mann-Whitney U test was employed to compare the mean survival fractions of cell lines transfected with siRNAs targeting an out-of-frame fusion gene partner with those of cell lines transfected with control siRNA. One-way ANOVA tests, with a Bonferroni correction for multiple testing, were used to compare the fold-change in cell population on the final day of each growth assay for cancer cells transfected with a fusion gene construct, full length 3' partner or empty vector. For all tests, two-tailed p values <0.05 were considered significant.

**REFERENCES**

1. Natrajan R, Weigelt B, Mackay A, *et al.* An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. *Breast Cancer Res Treat* 2010; **121**: 575-589.
2. Tognon C, Knezevich SR, Huntsman D, *et al.* Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell* 2002; **2**: 367-376.
3. Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: how special are they? *Mol Oncol* 2010; **4**: 192-208.
4. Weigelt B, Reis-Filho JS. Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol* 2009; **6**: 718-730.
5. Wetterskog D, Lopez-Garcia MA, Lambros MB, *et al.* Adenoid cystic carcinomas constitute a genomically distinct subgroup of triple-negative and basal-like breast cancers. *J Pathol* 2012; **226**: 84-96.
6. Marchio C, Iravani M, Natrajan R, *et al.* Mixed micropapillary-ductal carcinomas of the breast: a genomic and immunohistochemical analysis of morphologically distinct components. *J Pathol* 2009; **218**: 301-315.
7. Marchio C, Iravani M, Natrajan R, *et al.* Genomic and immunophenotypical characterization of pure micropapillary carcinomas of the breast. *J Pathol* 2008; **215**: 398-410.
8. Harvey JM, Clark GM, Osborne CK, *et al.* Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 1999; **17**: 1474-1481.
9. Jacobs TW, Gown AM, Yaziji H, *et al.* Specificity of HercepTest in determining HER-2/neu status of breast cancers using the United States Food and Drug Administration-approved scoring system. *J Clin Oncol* 1999; **17**: 1983-1987.
10. Marchio C, Natrajan R, Shiu KK, *et al.* The genomic profile of HER2-amplified breast cancers: the influence of ER status. *J Pathol* 2008; **216**: 399-407.
11. Mackay A, Tamber N, Fenwick K, *et al.* A high-resolution integrated analysis of genetic and expression profiles of breast cancer cell lines. *Breast Cancer Res Treat* 2009; **118**: 481-498.

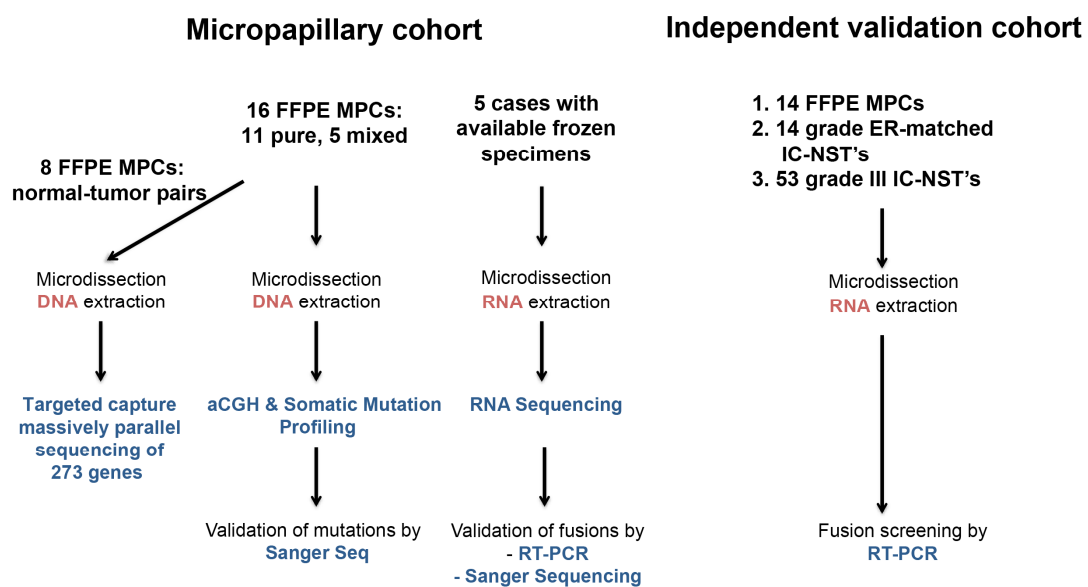


12. Duprez R, Wilkerson PM, Lacroix-Triki M, *et al.* Immunophenotypic and genomic characterization of papillary carcinomas of the breast. *J Pathol* 2012; **226**: 427-441.
13. Geyer FC, Weigelt B, Natrajan R, *et al.* Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. *J Pathol* 2010; **220**: 562-573.
14. Natrajan R, Lambros MB, Rodriguez-Pinilla SM, *et al.* Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res* 2009; **15**: 2711-2722.
15. Hernandez L, Wilkerson PM, Lambros MB, *et al.* Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. *J Pathol* 2012.
16. Wagle N, Berger MF, Davis MJ, *et al.* High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov* 2012; **2**: 82-93.
17. De Mattos-Arruda L, Bidard FC, Won HH, *et al.* Establishing the origin of metastatic deposits in the setting of multiple primary malignancies: The role of massively parallel sequencing. *Mol Oncol* 2013.
18. DePristo MA, Banks E, Poplin R, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491-498.
19. Cibulskis K, Lawrence MS, Carter SL, *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**: 213-219.
20. Ding J, Bashashati A, Roth A, *et al.* Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* 2012; **28**: 167-175.
21. McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-1303.
22. Koboldt DC, Zhang Q, Larson DE, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22**: 568-576.

23. Robinson JT, Thorvaldsdottir H, Winckler W, *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011; **29**: 24-26.
24. Sathirapongsasuti JF, Lee H, Horst BA, *et al.* Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* 2011; **27**: 2648-2654.
25. Ha KC, Lalonde E, Li L, *et al.* Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-mutated breast cancers and cell lines. *BMC Med Genomics* 2011; **4**: 75.
26. Banerji S, Cibulskis K, Rangel-Escareno C, *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 2012; **486**: 405-409.
27. Robinson DR, Kalyana-Sundaram S, Wu YM, *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat Med* 2011; **17**: 1646-1651.
28. Stephens PJ, McBride DJ, Lin ML, *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009; **462**: 1005-1010.
29. Network TCGA. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490**: 61-70.
30. Brough R, Frankum JR, Sims D, *et al.* Functional Viability Profiles of Breast Cancer. *Cancer Discov* 2011; **1**: 260-273.
31. Shiu KK, Wetterskog D, Mackay A, *et al.* Integrative molecular and functional profiling of ERBB2-amplified breast cancers identifies new genetic dependencies. *Oncogene* 2013.
32. Wilkerson PM, Dedes KJ, Wetterskog D, *et al.* Functional characterization of EMSY gene amplification in human cancers. *J Pathol* 2011; **225**: 29-42.
33. Dedes KJ, Wetterskog D, Mendes-Pereira AM, *et al.* PTEN deficiency in endometrioid endometrial adenocarcinomas predicts sensitivity to PARP inhibitors. *Sci Transl Med* 2010; **2**: 53ra75.

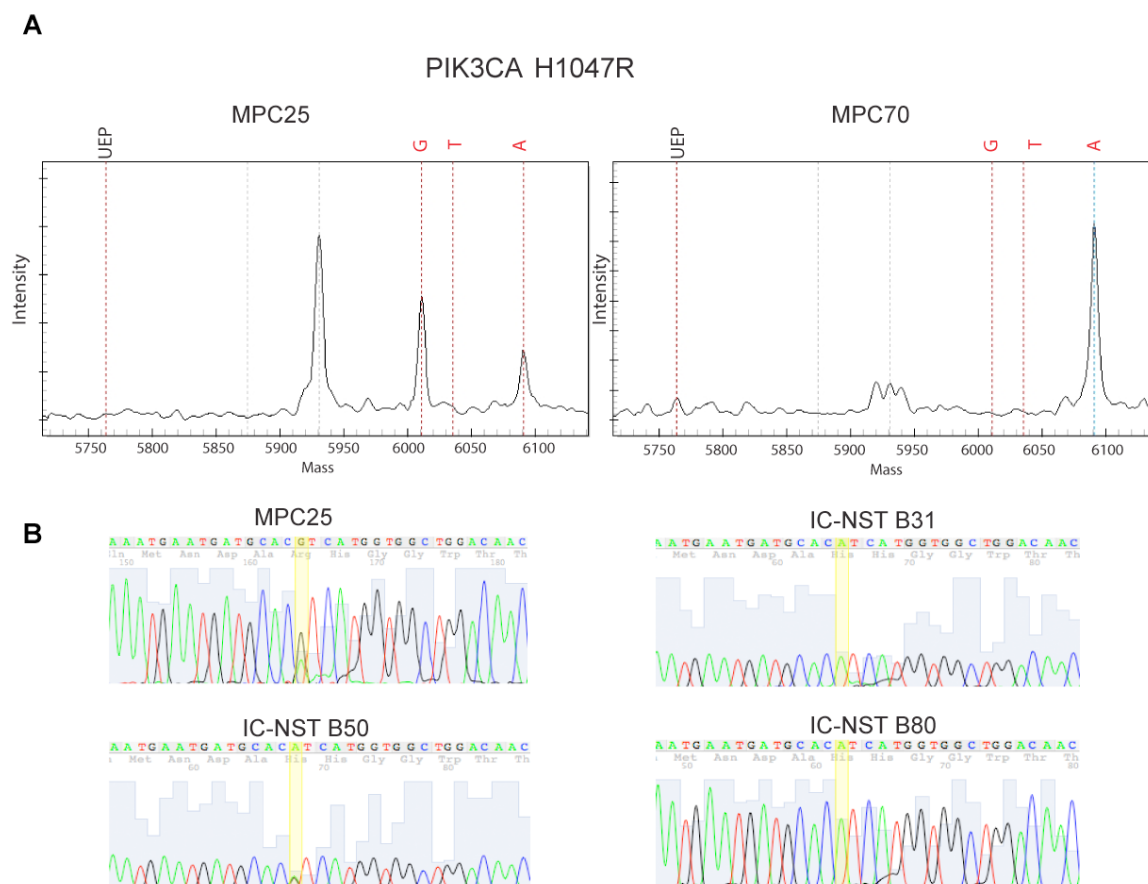
## SUPPLEMENTARY FIGURES

Supplementary Figure S1.



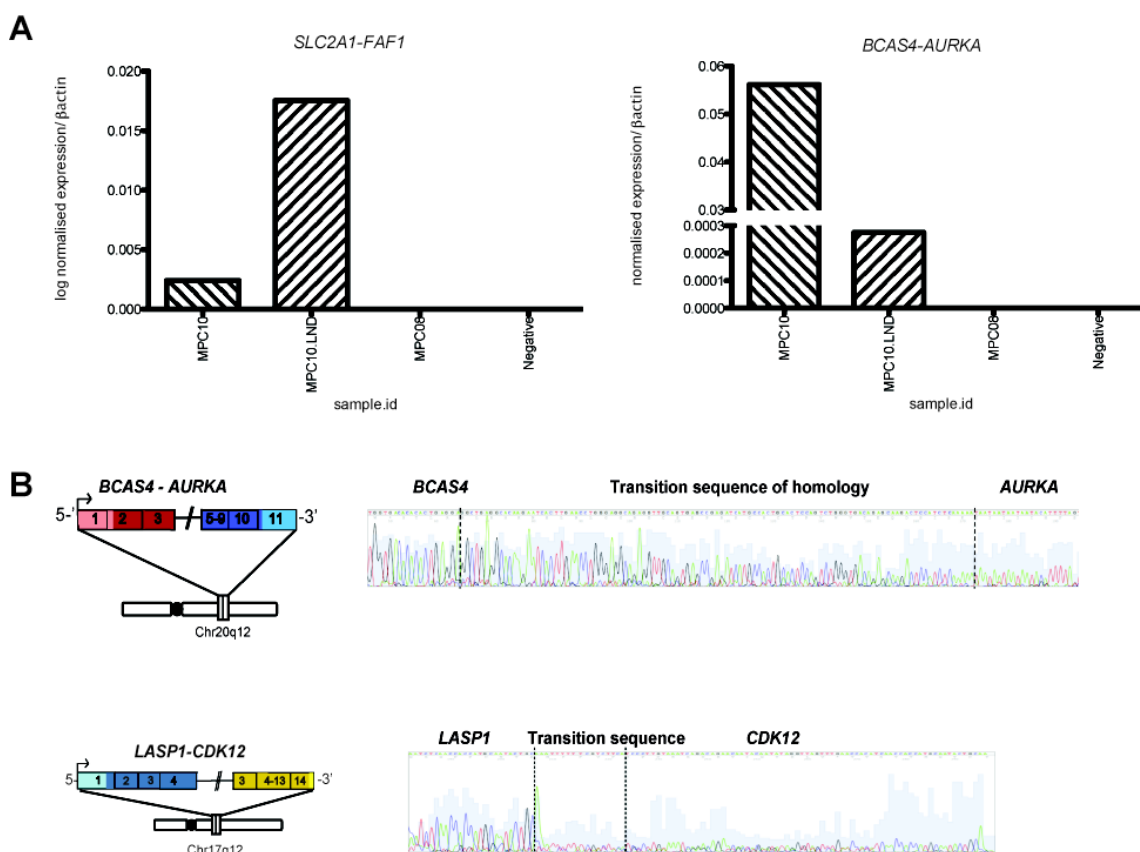
Supplementary Figure S1. Flow diagram of the study design.

## Supplementary Figure S2.



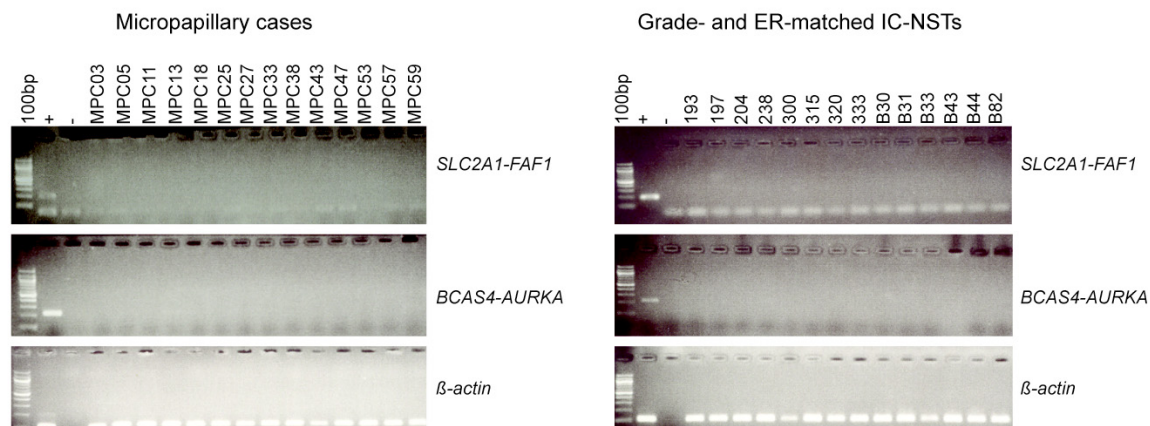
**Supplementary Figure S2. Identification and validation of *PIK3CA* mutations by the Sequenom Oncocarta v1.0 panel.** Using the Oncocarta v1.0 panel on the Sequenom platform, mutations were identified in a single micropapillary carcinoma (MPC25). **A**) Mass spectrometry profile illustrating the allele called in MPC25 and in the *PIK3CA* wild-type (WT) MPC70. **B**) Validation and identification of *PIK3CA* H1047R using conventional Sanger sequencing in MPC25 and in B50, the only case of the series of 16 grade-, ER-, and HER2-matched invasive ductal carcinomas of no special type (IC-NST) harboring a *PIK3CA* gene mutation. By comparison no mutations were identified in IC-NSTs B31 and B80.

## Supplementary Figure S3

**Supplementary Figure S3.** Validation of high-confidence fusion genes.

**A)** Real-time reverse transcription PCR (qRT-PCR) analysis of validated in-frame fusion genes in selected cases. qRT-PCR of *SLC2A1-FAF1* and *BCAS4-AURKA*. MPC.LND = lymph node metastasis of MPC10. Expression of fusion transcripts are normalized to the expression of  $\beta$ -actin. **B)** Long-range PCR validation of the DNA breakpoints of *BCAS4-AURKA*, *ELMO2-RAE1* and *LASP1-CDK12* in the index cases. Sanger sequencing validation of the DNA breakpoints of *BCAS4-AURKA* in MPC10 and *LASP1-CDK12* in MPC70. A schematic representation of nominated fusion transcripts is shown on the left. Fusion junctions with respective exon numbers are shown. Light colors represent 3' and 5' UTRs. Sequence chromatograms spanning the junction (dotted line) of the fusion transcript are shown on the right.

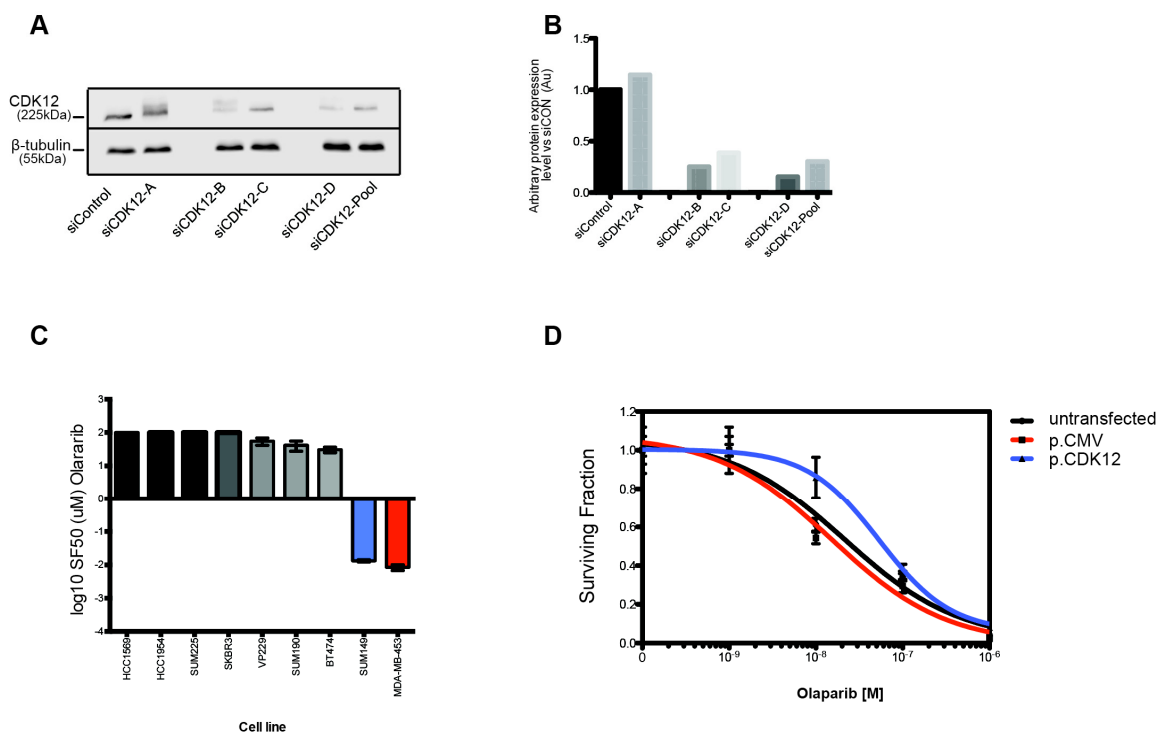
### Supplementary Figure S4



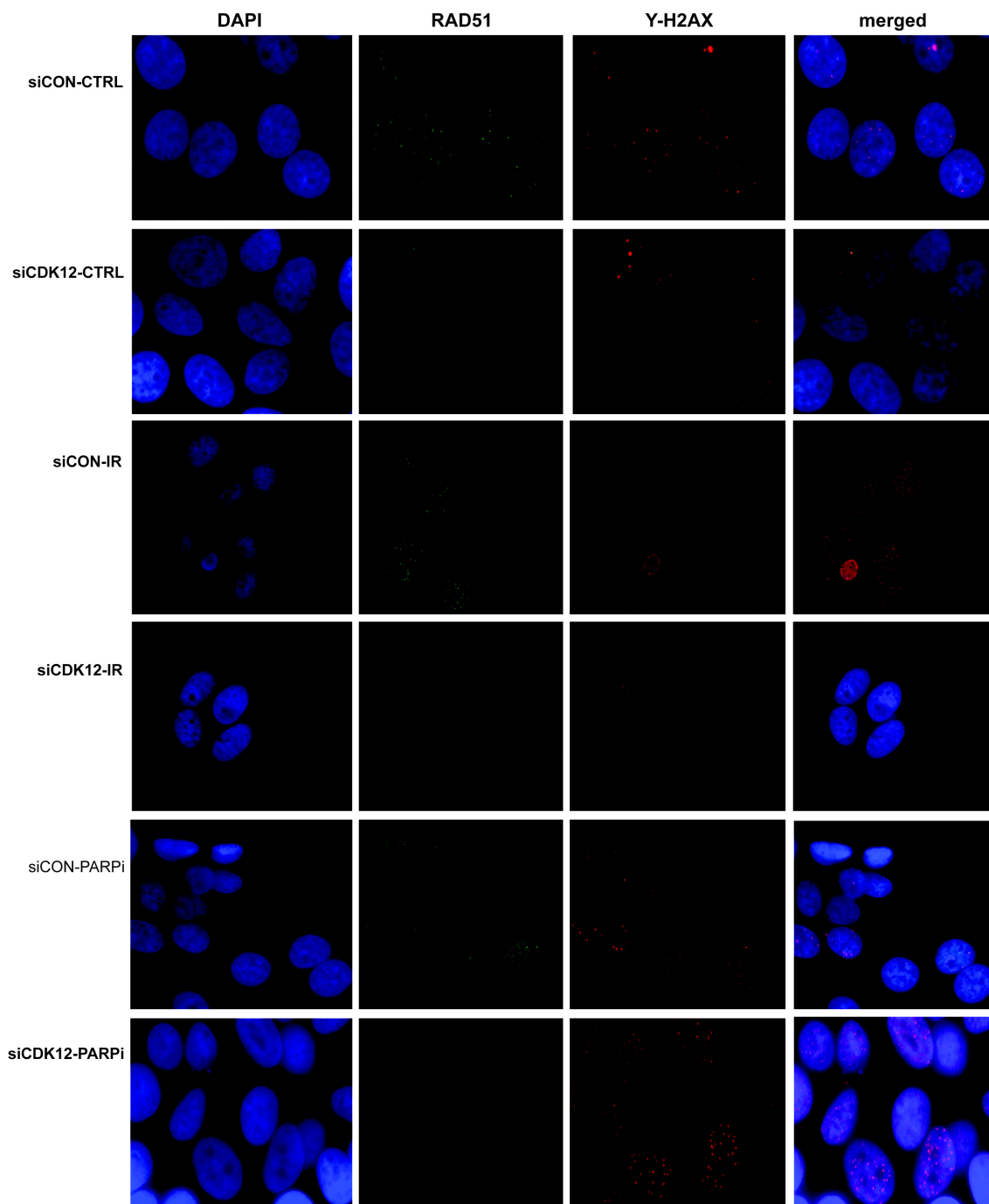
### Supplementary Figure S4. Recurrence RT-PCR screen for nominated fusions in 14 micropapillary and 14 grade- and ER-matched IC-NST.

To determine if the nominated fusion genes (in-frame fusions) were recurrent, 14 micropapillary (MPCs) and 14 grade- and ER-matched invasive carcinomas of no special type (IC-NSTs) were screened with primers for *SLC2A1-FAF1*, *BCAS4-AURKA*, and  $\beta$ -actin as a control. Positive (index case) and negative (no template) controls were included in each run.

## Supplementary Figure S5

**Supplementary Figure S5. Reconstitution of CDK12 in MDA-MB-453 with olaparib.**

**A)** Western blot analysis of siRNA against CDK12 and **B)** quantification, showing 3 out of 4 oligos and the pool reduce CDK12 protein expression. **C)** Bar plot of  $\log_{10}$  (SF50) of *HER2*-amplified cell lines to the PARP inhibitor olaparib. SUM149 (BRCA1 mutant) was used as a positive control (blue). MDA-MB-453 are sensitive to olaparib (red). **D)** Reconstitution of full length CDK12 (p.CDK12) in MDA-MB-453 results in increased resistance to olaparib (blue) relative to empty vector (red) (p.CMV) and untransfected cells (black).

**Supplementary Figure S6**

**Supplementary Figure S6. Assessment of RAD51 foci formation in MCF7 cells with CDK12 silencing.** Immunofluorescence images of representative fields of MCF7 cells with si-control (siCON) cells or CDK12-silenced (siCDK12) cells using a pool of four individual oligos showing DAPI-stained nuclei (blue),  $\gamma$ -H2AX (red), and RAD51 foci (green) after no treatment (CTRL), 10 Gy of irradiation (IR) or 24 hour treatment with 1 $\mu$ M of the PARP inhibitor BMN673.



Supplementary Table S1. Details of the complete cohort of breast cancers

Sample ID	Histology	Pure/ Mixed	Histological grade	ER	PR	HER2	aCGH	Sequenom	PIK3CA H1047R mutation screen	RNA-sequencing	RT-PCR fusion screen screen				
											Tested	BCAS4-AURKA	SLC2A1-FAF1	ELMO2-RAE1	LASP1-CDK12
MPC10	Micropapillary	pure	3	Positive	Positive	Negative	Y	Y	WT	Y	N	Y	Y	N	
MPC08	Micropapillary	pure	2	Positive	Positive	Negative	Y	Y	WT	Y	N	N	N	N	
MPC70	Micropapillary	pure	2	Positive	Positive	Negative	Y	Y	WT	Y	N	N	N	N	
MPC72	Micropapillary	pure	3	Positive	Negative	Negative	Y	Y	WT	Y	N	N	N	N	
MPC71	Micropapillary	pure	3	Positive	Positive	Positive	Y	Y	WT	Y	N	N	N	Y	
MPC37	Micropapillary	mixed	3	Positive	Positive	Negative	Y	Y	WT	N	N	NA	NA	NA	
MPC59	Micropapillary	mixed	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC27	Micropapillary	mixed	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC33	Micropapillary	mixed	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC03	Micropapillary	pure	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC13	Micropapillary	pure	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC43	Micropapillary	mixed	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC11	Micropapillary	pure	2	Positive	Positive	Negative	Y	Y	WT	N	Y	N	N	N	
MPC25	Micropapillary	mixed	2	Positive	Positive	Negative	Y	Y	H1047R	N	Y	N	N	N	
MPC57	Micropapillary	mixed	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC05	Micropapillary	pure	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC47	Micropapillary	mixed	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC18	Micropapillary	pure	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
MPC38	Micropapillary	mixed	3	Positive	Positive	Negative	Y	Y	WT	N	Y	N	N	N	
MPC53	Micropapillary	mixed	3	Negative	Negative	Positive	Y	Y	WT	N	Y	N	N	N	
MPC07	Micropapillary	pure	3	Positive	Positive	Negative	N	N	NA	N	N	NA	NA	NA	
MPC19	Micropapillary	pure	3	Positive	Positive	Negative	N	N	NA	N	N	NA	NA	NA	
MPC16	Micropapillary	pure	3	Positive	Positive	Negative	N	N	NA	N	N	NA	NA	NA	
MPC06	Micropapillary	pure	2	Positive	Positive	Negative	Y	Y	WT	N	N	NA	NA	NA	
MPC02	Micropapillary	pure	2	Positive	Positive	Negative	Y	Y	WT	N	N	NA	NA	NA	
MPC04	Micropapillary	pure	3	Positive	Negative	Negative	Y	Y	WT	N	N	NA	NA	NA	
MPC01	Micropapillary	pure	2	Positive	Positive	Negative	Y	Y	WT	N	N	NA	NA	NA	
MPC23	Micropapillary	pure	3	Positive	Negative	Negative	Y	Y	WT	N	N	NA	NA	NA	
MPC45	Micropapillary	mixed	3	Positive	Positive	Positive	Y	Y	WT	N	N	NA	NA	NA	
BC2468	IC-NST	NA	2	Positive	Positive	Negative	N	Y	WT	N	N	NA	NA	NA	
BC2472	IC-NST	NA	2	Positive	Positive	Negative	N	Y	WT	N	N	NA	NA	NA	
BC2522	IC-NST	NA	2	Positive	Positive	Negative	N	Y	WT	N	N	NA	NA	NA	
BC2512	IC-NST	NA	2	Positive	Negative	Negative	N	Y	WT	N	N	NA	NA	NA	
BC2469	IC-NST	NA	2	Positive	Positive	Negative	N	Y	WT	N	N	NA	NA	NA	
BC2473	IC-NST	NA	2	Positive	Positive	Negative	N	Y	WT	N	N	NA	NA	NA	
BC2700	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	N	NA	NA	NA	
B1	IC-NST	NA	3	Positive	Negative	Negative	N	Y	WT	N	Y	N	N	N	
B11	IC-NST	NA	3	Negative	Positive	Negative	N	N	NA	N	Y	N	N	N	
B12	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B15	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B16	IC-NST	NA	3	Positive	Positive	Negative	N	Y	WT	N	Y	N	N	N	
B17	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B2	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B20	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B21	IC-NST	NA	3	Positive	Positive	Negative	N	Y	WT	N	Y	N	N	N	
B24	IC-NST	NA	3	Positive	Negative	Negative	N	Y	WT	N	Y	N	N	N	
B26	IC-NST	NA	3	Positive	Negative	Positive	N	N	NA	N	Y	N	N	N	
B27	IC-NST	NA	3	Positive	Positive	Negative	N	Y	WT	N	Y	N	N	N	
B29	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B32	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B34	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B35	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B37	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B39	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B40	IC-NST	NA	3	Negative	Positive	Positive	N	N	NA	N	Y	N	N	N	
B41	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B45	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B46	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B47	IC-NST	NA	3	Positive	Negative	Negative	N	N	NA	N	Y	N	N	N	
B49	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B5	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B50	IC-NST	NA	3	Positive	Positive	Negative	N	Y	H1047R	N	N	NA	NA	NA	
B6	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B60	IC-NST	NA	3	Positive	Positive	Negative	N	Y	WT	N	Y	N	N	N	
B61	IC-NST	NA	3	Positive	Positive	Negative	N	Y	WT	N	Y	N	N	N	
B62	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B64	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B65	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B66	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B67	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B69	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B7	IC-NST	NA	3	Positive	Negative	Negative	N	N	NA	N	Y	N	N	N	
B72	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B74	IC-NST	NA	3	Positive	Negative	Negative	N	Y	WT	N	Y	N	N	N	
B76	IC-NST	NA	3	Positive	Positive	Positive	N	N	NA	N	Y	N	N	N	
B77	IC-NST	NA	3	Positive	Positive	Positive	N	N	NA	N	Y	N	N	N	
B78	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B79	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B80	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	
B81	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B84	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B86	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B87	IC-NST	NA	3	Negative	Negative	Negative	N	N	NA	N	Y	N	N	N	
B9	IC-NST	NA	3	Positive	Negative	Negative	N	Y	WT	N	Y	N	N	N	
193	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
197	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
204	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
238	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
315	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
320	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
333	IC-NST	NA	2	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
300	IC-NST	NA	2	Positive	Positive	Positive	N	N	NA	N	Y	N	N	N	
B30	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B31	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B33	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B44	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B82	IC-NST	NA	3	Positive	Positive	Negative	N	N	NA	N	Y	N	N	N	
B43	IC-NST	NA	3	Negative	Negative	Positive	N	N	NA	N	Y	N	N	N	

ER: estrogen receptor; PR: progesterone receptor; NA: not applicable; N: no; Y: yes.

**Supplementary Table S2.** Summary of antibodies, clones, dilutions, antigen retrieval methods, scoring and cut-offs used

Marker	Antibody clone	Dilution	Antigen retrieval	Company	Scoring system	Cut-off
EMA	E29	1:100	None	DAKO, Glostrup, Denmark	Qualitative assessment of the positivity: membranous or cytoplasmic; lining the luminal or stroma-facing border of the cells ('inside out' pattern)	-
ER	ID5	1:200	2 min, PC citrate, pH 6.0	DAKO, Glostrup, Denmark	Allred scoring system	Positive: >2 (ref 3)
PR	Pgr 636	1:200	3 min, PC citrate, pH 6.0	DAKO, Glostrup, Denmark	Allred scoring system	Positive: >2 (3)
HER2	Polyclonal	1:1200	41 min water bath, DAKO antigen retrieval solution, pH 6.0	DAKO, Glostrup, Denmark	Score 0, 1, 2, 3 according to Hercep Test kit scoring guidelines	Positive $\geq$ 2 (4)

PC=Pressure cooker

Supplementary Table S3: List of 273 genes included in the targeted capture massively parallel sequencing platform.

Gene_name	Number of cases with mutation in the TCGA breast cancer study	% of cases with mutation in the TCGA breast cancer study	Number of cases with amplification of homozygous deletion in the TCGA breast cancer study	% of cases with amplification of homozygous deletion in the TCGA breast cancer study
ABCA13	14	2.80%	2	0.40%
ABCB1	3	0.60%	1	0.20%
ADAMTSL1	3	0.60%	4	0.80%
AGFG2	3	0.60%	4	0.80%
AHNAK2	9	1.80%	6	1.20%
AK9	9	1.80%	2	0.40%
AKAP9	4	0.80%	16	3.20%
AKT1	12	2.40%	6	1.20%
AKT2	1	0.20%	7	1.40%
AKT3	3	0.60%	24	4.70%
ANK3	10	2.00%	9	1.80%
AOAH	0	0.00%	1	0.20%
APC	3	0.60%	0	0.00%
APOB	15	3.00%	0	0.00%
APOBEC1	1	0.20%	7	1.40%
APOBEC2	1	0.20%	5	1.00%
APOBEC3A	1	0.20%	1	0.20%
APOBEC3C	0	0.00%	1	0.20%
APOBEC3D	0	0.00%	2	0.40%
APOBEC3F	0	0.00%	2	0.40%
APOBEC3G	1	0.20%	2	0.40%
APOBEC3H	0	0.00%	2	0.40%
APOBEC4	1	0.20%	20	3.90%
ARAF	1	0.20%	0	0.00%
ARID1A	11	2.20%	2	0.40%
ATM	16	3.20%	5	1.00%
ATN1	8	1.60%	8	1.60%
ATR	2	0.40%	4	0.80%
ATRX	9	1.80%	0	0.00%
AURKA	1	0.20%	20	3.90%
AURKB	0	0.00%	2	0.40%
AURKC	1	0.20%	12	2.40%
BIRC5	0	0.00%	21	4.10%
BIRC6	6	1.20%	2	0.40%
BRAF	3	0.60%	3	0.60%
BRCA1	15	3.00%	5	1.00%
BRCA2	22	4.30%	8	1.60%
BRIP1	8	1.60%	36	7.10%
CACNA1A	6	1.20%	6	1.20%
CACNA1C	6	1.20%	7	1.40%
CACNA1E	13	2.60%	19	3.70%
CBFB	8	1.60%	8	1.60%
CDC25A	1	0.20%	0	0.00%
CDC25B	3	0.60%	5	1.00%
CDC25C	1	0.20%	1	0.20%
CDH1	33	6.50%	10	2.00%
CDK1	0	0.00%	9	1.80%
CDK4	0	0.00%	10	2.00%
CDK6	1	0.20%	3	0.60%
CDKN1A	0	0.00%	3	0.60%
CDKN1B	5	1.00%	2	0.40%
CDKN2A	0	0.00%	23	4.50%
CDKN2B	0	0.00%	24	4.70%
CEP164	4	0.80%	0	0.00%
CFH	8	1.60%	23	4.50%
CHD4	7	1.40%	8	1.60%
CHD6	7	1.40%	6	1.20%
CHEK1	0	0.00%	2	0.40%
CHEK2	4	0.80%	3	0.60%
COL12A1	10	2.00%	3	0.60%
CROCCP2	0	0.00%	0	0.00%
CSMD1	15	3.00%	23	4.50%
CSMD2	10	2.00%	0	0.00%
CSMD3	9	1.80%	57	11.20%
CTCF	13	2.60%	8	1.60%
CTNNB1	0	0.00%	1	0.20%
CUBN	7	1.40%	6	1.20%
DCHS2	8	1.60%	3	0.60%
DEPTOR	0	0.00%	56	11.00%
DMC1	2	0.40%	2	0.40%
DOCK11	9	1.80%	0	0.00%
DSPP	7	1.40%	4	0.80%
DST	7	1.40%	6	1.20%
EGFR	4	0.80%	8	1.60%
EME1	0	0.00%	34	6.70%

EME2	0	0.00%	12	2.40%
EPPK1	1	0.20%	35	6.90%
ERBB2	7	1.40%	66	13.00%
ERBB3	8	1.60%	2	0.40%
ERBB4	6	1.20%	3	0.60%
ERCC1	1	0.20%	6	1.20%
ERCC2	0	0.00%	6	1.20%
ERCC3	2	0.40%	1	0.20%
ERCC5	4	0.80%	8	1.60%
ESR1	2	0.40%	10	2.00%
ESR2	2	0.40%	4	0.80%
FAM157B	1	0.20%	2	0.40%
FAM186A	4	0.80%	1	0.20%
FANCA	7	1.40%	13	2.60%
FANCB	0	0.00%	0	0.00%
FANCC	1	0.20%	2	0.40%
FANCD2	3	0.60%	7	1.40%
FANCE	1	0.20%	1	0.20%
FANCF	1	0.20%	6	1.20%
FANCG	1	0.20%	4	0.80%
FANCI	3	0.60%	10	2.00%
FANCL	1	0.20%	3	0.60%
FANCM	2	0.40%	3	0.60%
FAT3	20	3.90%	8	1.60%
FBN1	7	1.40%	3	0.60%
FCGBP	6	1.20%	9	1.80%
FGFR1	0	0.00%	54	10.70%
FGFR2	4	0.80%	9	1.80%
FGFR3	1	0.20%	5	1.00%
FGFR4	3	0.60%	5	1.00%
FLG	22	4.30%	34	6.70%
FMN2	8	1.60%	24	4.70%
FOXA1	8	1.60%	8	1.60%
FOXC2	0	0.00%	12	2.40%
FRG1B	0	0.00%	2	0.40%
FRMPD4	7	1.40%	0	0.00%
GATA3	54	10.70%	12	2.40%
GATA4	0	0.00%	16	3.20%
GPR98	11	2.20%	2	0.40%
GRB2	1	0.20%	20	3.90%
GRIN2A	10	2.00%	10	2.00%
GRIN2B	8	1.60%	3	0.60%
HECW1	8	1.60%	1	0.20%
HERC2	6	1.20%	5	1.00%
HIF1A	1	0.20%	4	0.80%
HMCN1	19	3.70%	20	3.90%
HRAS	0	0.00%	4	0.80%
HRNR	9	1.80%	34	6.70%
HSP90AA1	2	0.40%	4	0.80%
HSP90AB1	2	0.40%	6	1.20%
HUWE1	9	1.80%	0	0.00%
HYDIN	13	2.60%	10	2.00%
IGF1R	2	0.40%	21	4.10%
INPP4B	4	0.80%	6	1.20%
IRS1	1	0.20%	1	0.20%
JAK1	2	0.40%	6	1.20%
JAK2	5	1.00%	4	0.80%
KIT	5	1.00%	0	0.00%
KRAS	4	0.80%	7	1.40%
LAMA1	7	1.40%	5	1.00%
LAMA5	7	1.40%	20	3.90%
LRP2	17	3.40%	2	0.40%
MACF1	11	2.20%	1	0.20%
MAP1A	10	2.00%	1	0.20%
MAP2K1	1	0.20%	4	0.80%
MAP2K2	1	0.20%	7	1.40%
MAP2K3	1	0.20%	7	1.40%
MAP2K4	21	4.10%	11	2.20%
MAP2K6	0	0.00%	23	4.50%
MAP3K1	39	7.70%	4	0.80%
MAP3K10	4	0.80%	7	1.40%
MAP3K4	4	0.80%	4	0.80%
MAP4K4	1	0.20%	3	0.60%
MAPK1	1	0.20%	3	0.60%
MAPK8	2	0.40%	1	0.20%
MAPK9	0	0.00%	3	0.60%
MDM2	2	0.40%	19	3.70%
MDN1	11	2.20%	4	0.80%
MED12	8	1.60%	0	0.00%
MET	3	0.60%	0	0.00%
MGAM	6	1.20%	3	0.60%

MGMT	2	0.40%	6	1.20%
MLH1	1	0.20%	0	0.00%
MLH3	3	0.60%	3	0.60%
MLL2	8	1.60%	2	0.40%
MLL3	37	7.30%	1	0.20%
MRE11A	1	0.20%	9	1.80%
MSH2	2	0.40%	2	0.40%
MSH3	2	0.40%	1	0.20%
MSH4	1	0.20%	2	0.40%
MSH5	1	0.20%	1	0.20%
MSH6	4	0.80%	2	0.40%
MST1L	5	1.00%	0	0.00%
MTOR	8	1.60%	2	0.40%
MUTYH	2	0.40%	2	0.40%
MXRA5	7	1.40%	0	0.00%
MYH8	5	1.00%	2	0.40%
NBEAL2	6	1.20%	2	0.40%
NBN	2	0.40%	37	7.30%
NBPF1	1	0.20%	0	0.00%
NCOA3	6	1.20%	12	2.40%
NCOR1	17	3.40%	4	0.80%
NCOR2	3	0.60%	5	1.00%
NEB	19	3.70%	0	0.00%
NF1	14	2.80%	12	2.40%
NF2	2	0.40%	3	0.60%
NR1H2	3	0.60%	10	2.00%
NRAS	0	0.00%	2	0.40%
OBSCN	16	3.20%	21	4.10%
PALB2	0	0.00%	9	1.80%
PARP1	1	0.20%	22	4.30%
PARP2	1	0.20%	1	0.20%
PARP3	2	0.40%	1	0.20%
PCLO	6	1.20%	1	0.20%
PCNXL2	9	1.80%	24	4.70%
PDGFRA	3	0.60%	1	0.20%
PDGFRB	2	0.40%	2	0.40%
PGR	2	0.40%	7	1.40%
PIK3CA	178	35.10%	22	4.30%
PIK3CB	4	0.80%	5	1.00%
PIK3R1	13	2.60%	2	0.40%
PKHD1L1	12	2.40%	53	10.50%
PLEC	7	1.40%	35	6.90%
PLK1	2	0.40%	9	1.80%
PLXNA4	8	1.60%	1	0.20%
PMS1	5	1.00%	3	0.60%
PMS2	1	0.20%	3	0.60%
POLB	1	0.20%	37	7.30%
POLD1	0	0.00%	11	2.20%
POLE	5	1.00%	1	0.20%
POLH	2	0.40%	5	1.00%
POLQ	6	1.20%	2	0.40%
PRKCA	0	0.00%	36	7.10%
PRKCB	3	0.60%	9	1.80%
PRKCD	0	0.00%	1	0.20%
PRKCG	2	0.40%	11	2.20%
PTCH1	6	1.20%	2	0.40%
PTEN	18	3.60%	9	1.80%
PTK2	3	0.60%	37	7.30%
RAD50	2	0.40%	0	0.00%
RAD51	1	0.20%	3	0.60%
RAD51B	0	0.00%	1	0.20%
RAD51C	2	0.40%	33	6.50%
RAD51D	1	0.20%	10	2.00%
RAD52	0	0.00%	13	2.60%
RAD54B	4	0.80%	38	7.50%
RAD54L	2	0.40%	3	0.60%
RAF1	1	0.20%	4	0.80%
RB1	9	1.80%	7	1.40%
RBBP8	0	0.00%	3	0.60%
RELN	13	2.60%	1	0.20%
RICTOR	3	0.60%	6	1.20%
RPS6KB1	0	0.00%	47	9.30%
RPTOR	2	0.40%	14	2.80%
RUNX1	17	3.40%	3	0.60%
SHC1	1	0.20%	32	6.30%
SHROOM4	5	1.00%	0	0.00%
SMO	2	0.40%	3	0.60%
SOS1	4	0.80%	2	0.40%
SPEN	13	2.60%	0	0.00%
SPRY1	0	0.00%	4	0.80%
SPTA1	14	2.80%	24	4.70%

SRCAP	9	1.80%	15	3.00%
SSPO	15	3.00%	5	1.00%
STAT1	1	0.20%	4	0.80%
STAT3	1	0.20%	10	2.00%
SVEP1	11	2.20%	3	0.60%
SYNE1	17	3.40%	11	2.20%
SYNE2	17	3.40%	4	0.80%
TBX3	13	2.60%	1	0.20%
TENM1	10	2.00%	0	0.00%
TEX15	10	2.00%	14	2.80%
TGFBR1	2	0.40%	3	0.60%
TGFBR2	3	0.60%	0	0.00%
TGFBR3	2	0.40%	2	0.40%
TMC2	4	0.80%	2	0.40%
TOP2A	1	0.20%	24	4.70%
TP53	187	36.90%	3	0.60%
TP53BP1	3	0.60%	1	0.20%
TSC1	3	0.60%	2	0.40%
TSC2	2	0.40%	12	2.40%
TYK2	4	0.80%	6	1.20%
UBR4	12	2.40%	0	0.00%
USH2A	27	5.30%	21	4.10%
USP36	3	0.60%	19	3.70%
VWF	8	1.60%	7	1.40%
WDFY3	9	1.80%	3	0.60%
XBP1	2	0.40%	3	0.60%
XIRP2	12	2.40%	0	0.00%
XPA	0	0.00%	1	0.20%
XPC	0	0.00%	5	1.00%
XRCC1	1	0.20%	4	0.80%
XRCC2	0	0.00%	1	0.20%
XRCC3	0	0.00%	6	1.20%
ZFHX3	10	2.00%	9	1.80%
ZFHX4	16	3.20%	28	5.50%
ZNF384	0	0.00%	8	1.60%
ZNF703	0	0.00%	61	12.00%

Supplementary Table S4. Breast cancer cell line characteristics

	MCF7	BT474	T47D	ZR75.1	MDA-MB-361	HCC1954	MDA-MB-453	JIMT1	UACC812	UACC893	VP229	SKBR3	ZR75.30	SUM225	SUM190	HCC1569	SUM149
<b>ER</b>	Positive	Positive	Positive	Positive	Positive	Negative	Negative	Negative	Positive	Positive	Positive	Negative	Positive	Negative	Negative	Negative	Negative
<b>HER2</b>	Negative	Positive	Negative	Negative	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Positive	Negative
<b>TP53</b>	Negative	Positive/ Mutant	Positive/ Mutant	Negative	Negative/ WT	Negative/ Mutant	Negative/ WT	NA	Negative/ WT	Mutant	Mutant	Positive/ Mutant	Negative/ WT	Positive/ Mutant	Negative/ Mutant	Negative/ Mutant	Positive/ Mutant
<b>1q gain</b>	Gain	Gain	Gain	Gain	NC	Gain	Gain	Gain	Gain	Gain	NC	Gain	Gain	Gain	Gain	Gain	Gain
<b>8q gain</b>	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain
<b>17q gain</b>	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	NC	NC
<b>20q gain</b>	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	Gain	NC	Gain	NC	Gain	Gain
<b>1p loss</b>	Loss	Loss	Loss	Loss	Loss	Loss	NC	Loss	Loss	Loss	Loss	Loss	Loss	NC	NC	NC	NC
<b>8p loss</b>	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss	Loss
<b>13q loss</b>	Loss	NA	Loss	NA	NC	NC	Loss	NC	Loss	Loss	Loss	NC	NC	NC	Loss	Loss	NC
<b>16q loss</b>	Loss	Loss	Loss	Loss	NC	Gain	Loss	Loss	Loss	Loss	NC	Loss	Loss	NC	NC	NC	Loss
<b>22q loss</b>	Loss	Loss	Loss	Loss	Loss	Loss	Gain	Loss	Loss	Loss	NC	Loss	NC	Loss	Loss	NC	Loss
<b>RAE1</b>	Amp	Amp/fusion	NC	NC	NC	Amp	NC	Amp	Amp	NC	Amp	Amp	NC	NC	NC	NC	Amp
<b>ELMO2</b>	NC	NC	NC	NC	NC	NC	NC	Amp	NC	NC	Amp	Amp	NC	NC	NC	Amp	Amp
<b>BCAS4</b>	Amp/fusion	Amp	NC	NC	NC	NC	NC	Amp	Amp	NC	Amp	NC	NC	NC	NC	NC	NC
<b>AURKA</b>	Amp	Amp	NC	NC	NC	Amp	NC	Amp	Amp	NC	Amp	Amp	NC	NC	NC	NC	Amp
<b>SLC2A1</b>	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
<b>FAF1</b>	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
<b>CDK12</b>	NC	Amp	NC	NC	Fusion /breakpoint	Fusion/ breakpoint	Breakpoint	No breakpoint	No breakpoint	No breakpoint	No breakpoint	No breakpoint	No breakpoint	Breakpoint	Breakpoint	Breakpoint	NC
<b>LASP1</b>	NC	Amp	NC	NC													
<b>BRCA1</b>	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	Mutant

Amp: amplification; fusion: fusion gene detected involving this gene; gain: chromosomal arm gain; loss: chromosomal arm loss; NC: no copy number change/alteration, not assessed breakpoint; breakpoint identified in gene from aCGH analysis.







Main table containing genomic coordinates, gene names, and associated data points for various genomic regions.

LOSSES (3 or more cases)

Table with columns for Chr, Start, End, Start MB, End MB, Cytobands, Start Prob, End Prob, Probe, Length MB, max, coverage, and Cases, listing genomic losses and associated gene names.

Table with columns for Chr, Start, End, Start MB, End MB, Cytobands, Start Prob, End Prob, Probe, Length MB, max, coverage, and Cases, listing genomic losses and associated gene names.

Table with columns: ID, Date, Time, Location, Name, and Description. Contains a large list of entries with alphanumeric codes and names.





**Supplementary Table S6.** Summary of mRNA-sequencing statistics

<b>Sample ID</b>	<b>Total sequencing per lane (GB)</b>	<b>Aligned reads per lane (GB)</b>
MPC08	1.43	1.77
MPC10	1.92	2.67
MPC70	1.33	1.89
MPC71	1.37	1.88
MPC72	1.35	1.83

**Supplementary Table S7.** Recurrent alterations identified through re-analysis of public datasets (n=185)

Index case			Recurrence					
Fusion partner gene	Index case	Index case in frame	Sample characteristics	Sample phenotype	In frame	DNA or RNA	Reference	Number of samples sequenced in study
BCAS4	MPC10	Y	Cell line MCF7	ER+ HER2-	Y	RNA	Robinson et al; Edgren et al	89; 5
C17orf57	MPC71	N	Primary breast cancer	ER-HER2+	N	DNA	Nik-Zainal et al	21
CDK12	MPC71	N	Cell line HCC1954	ER- HER2+	Unknown	DNA	Stephens et al	24
CDK12	MPC71	N	Cell line HCC2218	ER- HER2+	Unknown	DNA	Stephens et al	24
CDK12	MPC71	N	Primary breast cancer	ER- HER2+	N	DNA	Nik-Zainal et al	21
CDK12	MPC71	N	Cell line MDA-MB-361	ER+ HER2+	Y	RNA	Robinson et al	89
CDK12	MPC71	N	Primary breast cancer	HER2+	Y	DNA	Banerji et al	22
FAF1	MPC10	N	Primary breast cancer	ER- HER2-	N	DNA	Nik-Zainal et al	21
LASP1	MPC71	N	Primary breast cancer	ER+ HER2-	N	DNA	Nik-Zainal et al	21
LASP1	MPC71	N	Primary breast cancer	Unknown	Unknown	DNA	Stephens et al	24
NSF	MPC71	N	Primary breast cancer	ER+ HER2+	Unknown	DNA	Nik-Zainal et al	21
RAE1	MPC10	N	Cell line BT474	ER+ HER2+	N	RNA	Robinson et al	89
RAE1	MPC10	N	Primary breast cancer	ER+ HER2-	N	RNA	Robinson et al	89
USH2A	MPC10	N	Primary breast cancer	ER+ HER2-	N	DNA	Nik-Zainal et al	21
USH2A	MPC10	N	Primary breast cancer	Unknown	N	DNA	Ellis et al	46