CLINICAL TRIAL

# Mutation profiling identifies numerous rare drug targets and distinct mutation patterns in different clinical subtypes of breast cancers

**Libero Santarpia · Yuan Qi · Katherine Stemke-Hale · Bailiang Wang · Elliana J. Young · Daniel J. Booser · Frankie A. Holmes · Joyce O'Shaughnessy · Beth Hellerstedt · John Pippen · Tatiana Vidaurre · Henry Gomez · Vicente Valero · Gabriel N. Hortobagyi · W. Fraser Symmans · Giulia Bottai · Angelo Di Leo · Ana M. Gonzalez-Angulo · Lajos Pusztai**

**Abstract** The mutation pattern of breast cancer molecular subtypes is incompletely understood. The purpose of this study was to identify mutations in genes that may be targeted with currently available investigational drugs in the three major breast cancer subtypes (ER+/HER2−, HER2+, and Triple Negative). We extracted DNA from fine needle aspirations of 267 stage I–III breast cancers. These tumor specimens typically consisted of >80 % neoplastic cells. We examined 28 genes for 163 known cancer-related nucleic acid variations by Sequenom technology. We observed at least one mutation in 38 alleles corresponding to 15 genes in 108 (40 %) samples, including *PIK3CA* (16.1 % of all samples), *FBXW7* (8 %), *BRAF* (3.0 %), *EGFR* (2.6 %), *AKT1* and *CTNNB1* (1.9 % each), *KIT* and *KRAS* (1.5 % each), and *PDGFR-α* (1.1 %). We also checked for the polymorphism in *PHLPP2* that is known to activate *AKT* and it was found at 13.5 % of the patient samples. *PIK3CA* mutations were more frequent in estrogen receptor-positive cancers compared to triple negative breast cancer (TNBC) (19 *vs.* 8 %, *p = 0.001*). High frequency of *PIK3CA* mutations (28 %) were also found in HER2+ breast tumors. In TNBC, *FBXW7* mutations were significantly more frequent compared to ER+ tumors (13 *vs.* 5 %, *p = 0.037*). We performed validation for all mutated alleles with allele-specific PCR or direct sequencing; alleles analyzed by two different sequencing techniques showed 95–100 % concordance for mutation status. In conclusion, different breast cancer subtypes harbor different type of mutations and approximately 40 % of tumors contained individually rare mutations in

L. Santarpia (✉) · G. Bottai · A. Di Leo
Department of Oncology, Translational Research Unit, Hospital of Prato and Istituto Toscano Tumori, 59100 Prato, Italy
e-mail: lsantarpia@usl4.toscana.it

L. Santarpia · W. F. Symmans
Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77230-1439, USA

L. Santarpia · B. Wang · E. J. Young · D. J. Booser · V. Valero · G. N. Hortobagyi · A. M. Gonzalez-Angulo · L. Pusztai (✉)
Department of Breast Medical Oncology, Unit 1354, The University of Texas MD Anderson Cancer Center, PO Box 301439, Houston, TX 77230-1439, USA
e-mail: lpusztai@mdanderson.org

Y. Qi
Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77230-1439, USA

K. Stemke-Hale · A. M. Gonzalez-Angulo
Department of System Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77230-1439, USA

F. A. Holmes · J. O'Shaughnessy · J. Pippen
Texas Oncology/US Oncology, Houston, TX 75246, USA

J. O'Shaughnessy · J. Pippen
Baylor Sammons Cancer Center, Texas Oncology, US Oncology, Dallas, TX 75246, USA

B. Hellerstedt
Texas Oncology/US Oncology, Austin, TX 78745, USA

T. Vidaurre · H. Gomez
Department of Oncological Medicine, Instituto Nacional de Enfermedades Neoplasicas, Lima 34, Peru

signaling pathways that can be potentially targeted with drugs. Simultaneous testing of many different mutations in a single needle biopsy is feasible and allows the design of prospective clinical trials that could test the functional importance of these mutations in the future.

**Keywords** Breast cancer · Gene mutations · Signaling pathways · Mutation detection · Drug targets

## Introduction

All cancers carry a complex repertoire of nucleic acid variations including germ line polymorphisms and somatic mutations (http://www.sanger.ac.uk/genetics/CGP/Census) that can influence the clinical behavior of the disease and may represent potential therapeutic targets [1, 2]. The mutation landscape of various types of cancers is currently being mapped with the expectation that some of the genomic alterations will turn out to be important therapeutic targets [3]. Small scale, early breast cancer studies that used whole or partial genome sequencing, revealed mutations in dozens of potentially cancer-causing genes that often appeared to be unique to a given case [4–9]. The number of patients included in these studies ranged from 1 to 16. The two largest mutation profiling studies in breast cancer tested around 30 oncogenes and included 183 and 53 cases, respectively, therefore the true prevalence and pattern of mutations in breast cancer subtypes remains uncertain [10, 11]. Also, the aforementioned studies provided no clinical and pathological correlations or examined association between mutations and response to treatment.

The purpose of the current study was to assess the prevalence of a selected number of somatic mutations and functional SNP in 267 stage I–III breast cancer samples with known clinical outcome. These genes and nucleic variations were selected because of their known functional importance in cancer biology and because they represented molecules or signaling pathways that can be targeted with approved or investigational drugs currently in clinical trials. We analyzed fine needle aspirations (FNA) of stage I–III breast cancers collected before neoadjuvant chemotherapy in the context of two prospective and IRB-approved clinical trials. Importantly, FNA specimens are highly enriched in neoplastic cells with a few contaminating leukocytes and minimal or no stromal (fibroblast, adipocyte, and endothelial cell) contamination; therefore, our results likely represent mutational status of the neoplastic cells in these cancers [12]. Mutation analysis was performed using Sequenom technology [13]. There were no normal tissues included in this analysis, somatic nucleic acid variations were retrieved from the public dbSNP database for comparison.

## Materials and methods

### Tumor specimens and nucleic acids extraction

FNA samples of newly diagnosed stage I–III breast cancers were obtained with a 23-gauge needle, and cells from two passes were collected into a vial containing 1 ml of RNAlater solution (Ambion, Austin, Texas). The vials were stored at −80 °C until total RNA and DNA extraction. One hundred seventy-five biopsies were collected at The University of Texas M.D. Anderson Cancer Center (MDACC, Houston, Texas), 76 specimens were received from US Oncology, Inc. (USO, Houston, Texas) and 16 specimens were collected at the Instituto Nacional de Enfermedades Neoplásicas (INEN, Lima, Peru). Results of these clinical trials were reported separately [14, 15]. Patient characteristics are presented in Table 1.

All clinical information including survival and recurrence dates were extracted from the prospectively collected and maintained clinical trial information databases at MDACC (search date: June 28, 2010) and USO (search date: July 7, 2010), respectively. Pathologic complete response (pCR) was defined as no residual invasive cancer in the breast or lymph nodes. All other cases were categorized as residual cancer (RD). Recurrence-free survival estimates were calculated from the date of the diagnostic biopsy until first recurrence. Overall survival was calculated from the date of the diagnostic biopsy until the date of death. Clinical ER and HER2 status were determined in a diagnostic core biopsy by the routine clinical pathology laboratories at the site of the biopsy collection. ER positivity was defined as nuclear staining in $\geq 10$ % of cancer cells and HER-2 overexpression was defined as a gene copy ratio of HER-2 gene:chromosome 17 centromere $\geq 2.0$ or an IHC score of 3+. The HER-2 status of tumors with an IHC score of 1+ or 2+ were confirmed using FISH.

DNA was extracted from the flow-through specimens of a preceding RNA extraction step that used RNeasy mini kits (Qiagen, Valencia, California). DNA extraction was performed using the QIAamp DNA Mini Kit (Qiagen) following the manufactures instructions. Final DNA concentration was assessed using Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc, Pittsburgh, Pennsylvania).

### Detection of mutations and single nucleotide polymorphisms

We analyzed 163 somatic mutations and single nucleotide polymorphisms (SNP) in 28 cancer-related genes (Supplementary Table S1) in 267 stage I–III breast cancer samples. These nucleic acid variations were selected from

**Table 1** Patients' characteristics at diagnosis

| Total no. of patients | | 267 | % |
|---|---|---|---|
| Age | Mean (mini-max) | 50 (24–80) | |
| Tumor size | | | |
| | T1 | 6 | 2, 2 |
| | T2 | 107 | 40, 1 |
| | T3 | 61 | 22, 8 |
| | T4 | 93 | 34, 8 |
| Lymph node status | | | |
| | Node-positve | 191 | 71, 5 |
| | Node-negative | 73 | 27, 3 |
| | Unknown | 3 | 1, 1 |
| Histologic grade | | | |
| | I | 5 | 1, 9 |
| | II | 84 | 31, 5 |
| | III | 159 | 59, 6 |
| | Unknown | 19 | 7, 1 |
| Estrogen receptor (ER) status[a] | | | |
| | ER-positive | 111 | 41, 6 |
| | ER-negative | 155 | 58, 1 |
| | Unknown | 1 | 0, 4 |
| HER-2 status[a] | | | |
| | HER-2-positive | 61 | 22, 8 |
| | HER-2-normal | 204 | 76, 4 |
| | Unknown | 2 | 0, 7 |
| Neoadjuvant chemotherapy[b] | | | |
| | FAC or FEC | 49 | 18, 4 |
| | T/FAC | 63 | 23, 6 |
| | TX/FAC | 120 | 44, 9 |
| | Other | 29 | 10, 9 |
| | None | 6 | 2, 2 |
| Pathologic response[c] | | | |
| | pCR | 67 | 25, 1 |
| | No-pCR | 180 | 67, 4 |
| | Unknown (no surgery or lost f/u) | 20 | 7, 5 |
| Recurrence (local or distance) | | | |
| | Recurrence | 66 | 24, 7 |
| | No recurrence | 196 | 73, 4 |
| | Unknown | 5 | 1, 9 |
| Survival | | | |
| | Deceased | 61 | 22, 8 |
| | Alive | 203 | 76 |
| | Unknown | 3 | 1, 1 |

[a] Receptor status was assessed by the pathology lab as routine care

[b] FAC = 5-fluorouracil, doxorubicin (or $E$ epirubicin) and cyclophosphamide every 3 weeks × 4, T/FAC = weekly paclitaxel × 12 followed by FAC. TX/FAC = docetaxel × 4 treatments followed by FAC × 4

[c] pCR refers to pathologic complete response in the breast and axilla

public databases including COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/) and the NCI somatic mutation data bases http://tcga-data.nci.nih.gov/docs/somatic_mutations/tcga_mutations.htm). We based our selection criteria on nonsynonymous coding mutations that were reported to occur in human cancers and are considered to be functionally relevant as well as potentially druggable with currently available investigational already approved drugs.

We used Sequenom technology to identify known single nucleotide variations. DNA regions flanking the selected mutation or SNP sites were amplified by PCR in a 5 μl multiplex reaction using Qiagen Hotstar Mastermix supplemented with 5 mM $MgCl_2$ (Qiagen, Valencia,

California). Unincorporated primers and deoxyribonucleotide triphosphates (dNTPs) were removed using ExoSAP available as part of the Sequenom iPLEX Gold kit (Sequenom, San Diego, California). Primer extension reactions were performed using the iPLEX Gold Taq, and the sample was desalted using Clean Resin (Sequenom). PCR and extensions primers were designed by the Sequenom Assay Design Software then checked for false priming and common SNP sites with proprietary software available on http://www.mysequenom.com web site. The desalted primer extension reactions were spotted onto SpectroChip II and run in the Sequenom MassARRAY. All spectra corresponding to sequence data were generated in duplicate on the same run and were interpreted by the Typer Software 3.4 and 4.0. In-house software was used to compare duplicate and to calculate the relative amount of each base. Reactions with >15 % of the detected mass corresponding to the mutated allele were called positive for a nucleic acid variation. When duplicate cells were discordant, the spectra were visually inspected, if the quality of one spectrum was bad, the call was based on the single good spectrum; if both spectra were good but discrepant, the call was discarded from further analysis.

### Allelic discrimination by quantitative real- time polymerase chain reaction (PCR) and automatic sequencing

To validate mutations identified by Sequenom and to avoid false positive call, we performed allelic discrimination qRT-PCR (i.e., EGFR, KIT, and PHLPP2) or direct sequencing (for all the other genes) for all mutant alleles. Polymorphisms in DNA sequence were assessed by the 5′ nuclease allelic discrimination assay or by direct sequencing. For allelic discrimination PCR, the region flanking the polymorphism was amplified by PCR in the presence of two fluorescent probes, each specific for one allele. The primers were designed using Primer Express 3.0 (Applied Biosystems, Foster City, California) with attention to avoid SNPs in the flanking region of the mutation. Each PCR mixture (12.5 μl) contained 20 ng DNA, 900 nM of forward and reverse primer each, 300 nM of allele specific probes each, and 6.25 μl of TaqMan Universal PCR Master Mix (Applied Biosystems). Amplification was done under the following conditions: 50 °C for 2 min, 95 °C for 10 min followed by 40 cycles of 92 °C for 15 s, and 60 °C for 1 min. Fluorescence in each sample well was measured before and after PCR using ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Data were analyzed using Allelic Discrimination Program (Applied Biosystems).

Automatic sequencing was performed using an ABI 3730xl DNA Analyzer (Applied Biosystems). For each mutation, DNA samples were genotyped twice to confirm the results. Data were analyzed by Applied Biosystems Seqscape v2.6 software, which defines a heterozygous mutation as a minor variant peak ≥25 % of the intensity of the dominant peak.

### Statistical analysis

Individual nucleic acid variations in a given gene were collapsed into gene level mutation status, and a gene was considered mutated if it had any nucleic acid variation. Genes with mutations in at least one sample were also collapsed into three functional pathway groups including: (i) phosphatidylinositol-3-kinases/v-akt thymoma viral oncogene homolog 1 (PI3K) pathway (AKT-1,-2,-3, FRAP, PDK1, PHLPP2, PIK3CA, PI3KR1, RICTOR, and PRKAG-1), (ii) receptor tyrosine kinase (RTK) group (EGFR, FGFR-1, KIT, IGF1R, MET, RET, and PDGFR-α), and (iii) Cell cycle/metabolic pathway group (BRAF, CTNNB1, DEAR1, FBXW7, HIF1-α, IDH-1, -2, JAK2, KRAS, and TNK2). At subsequent analysis, a mutation or the presence of the activating polymorphism in PHLPP2 in any one of the member genes qualified a patient for inclusion in that mutated pathway category. Mutation frequencies at allele, gene, and pathway levels were compared across three clinical subsets of breast cancers including: (i) ER-positive/HER2-normal (n = 88), (ii) HER2-positive with any ER status (n = 61), and (iii) ER-negative/HER2-negative (triple negative, TNBC, n = 116) using the Fisher's Exact Test (FET). Univariate and multivariate logistic regression analysis were also performed. The predictor variables of the regression model were clinical ER status (positive vs. negative), HER2 status (positive vs. negative), tumor grade (level 2/3 vs. level 1), nodal status (positive vs. negative), and pathologic response to preoperative chemotherapy (pCR vs. RD). Survival by individual gene mutation status and by pathway mutation status was plotted using the Kaplan–Meier method. Survival curves were plotted separately for each disease subsets to avoid the confounding effect of mutation status associated with disease subtype.

## Results

### Distinct mutation patterns in different breast cancer subtypes

We detected mutations in 38 alleles in 15 genes in 108 (40 %) samples. Table 2 shows each of the detected mutations and their frequencies along with the potential therapeutic agents in clinical trials that could target these genes.

**Table 2** Observed mutations in breast cancer and corresponding drugs that inhibit functional signaling pathways

| Gene symbol | Genomic and codon positions, base and amino acid substitutions* | Number (%) of cases with mutations at allele level | Number (%) of cases with mutation at gene level | Drugs that inhibit gene function or block the involved signaling pathway |
|---|---|---|---|---|
| AKT1 | c.49G > A/p.E17K[1] | 5 (1.9 %) | 5 (1.9 %) | MK2206, GSK690693, KRX-041 (Perifosine), GSK2141795 |
| BRAF | c.1391G > T/p.G464V[1] | 1 (0.4 %) | 8 (3 %) | Sorafenib Tosylate (Nexavar, or Bay 43-9006), CI-1040 (PD184352), PLX4720, GDC0879, AZD6244 (Selumetinib), PD98059, SL327, PD0325901, BIRB796 (Doramapimod), SD169, SB202190, SB203580, PD169316, PLX4032, AS703026 |
| | c.1397G > T/p.G466V[1] | 1 (0.4 %) | | |
| | c.1396G > C/p.G466R[1] | 4 (1.5 %) | | |
| | c.1798G > A/p.V600M[1] | 1 (0.4 %) | | |
| | c.1799T > A/p.V600E[1] | 1 (0.4 %) | | |
| CTNNB1 | c.101G > T/p.G34V[1] | 1 (0.4 %) | 5 (1.9 %) | AVN316, inhibitors of beta catenin—Wnt signaling |
| | c.121A > G/p.T41A[1] | 4 (1.5 %) | | |
| EGFR | c.2573T > G/p.L858R[1] | 7 (2.6 %) | 7 (2.6 %) | Erlotinib Hydrochloride (Tarceva, CP-358774, OSI-774, NSC 718781), Lapatinib Ditosylate (Tykerb, Tyverb, o GW-572016), Gefitinib (Iressa o ZD-1839), Cetuximab (IMC-C225 o Erbitux), CI-1033 (Canertinib, PD-183805, CII033, o PD183805), ZD6474 (Vandetanib), BIBW-2992 (Afatinib, INN, Tovok o Tomtovok), XL647, CUDC101, AG-1478 (NSC 693255, Tyrphostin AG-1478), PD153035 hydrochloride, AG-18 (RG-50810, Tyrphostin AG-18, Tyrphostin 23, TX 825, RG-50858), AG-213 (Tyrphostin AG-213), NVP-AEE788 |
| FBXW7 | c.1436G > T/p.R479I[2] | 1 (0.4 %) | 21 (7.9 %) | Tumor cell with FBXW7 mutation are particularly sensitive to inhibitors of the mTOR pathway |
| | c.1514G > A p.R505H[2] or c.1514G > T/p.R505L[2] or c.1514G > C/p.R505P[2] | 14 (5.1 %) | | |
| | c.1745C > T/p.S582L[2] | 4 (1.5 %) | | |
| | c.1379A > G/p.H460R[2] | 2 (0.7 %) | | |
| HIF1-α | c.2089C > G/p.Q697E[3] | 1 (0.4 %) | 1 (0.4 %) | PX478, RX0047, SF1126, 2 methoxyestradiol (2-ME2) |
| IDH2 | c.515G > A/p.R172K[1] | 1 (0.4 %) | 1 (0.4 %) | Gliomas with IDH2 mutation may have higher sensitivity to temazolamide |
| KIT | c.2446G > C/p.D816H[1] | 1 (0.4 %) | 4 (1.5 %) | Imatinib mesylate (Gleevec), Nilotinib (AMN107), Axitinib (AG-013736), XL184, Masitinib mesylate, Dasatinib (BMS-354825, Sprycel, o BMS354825, AZD0530 (Saracatinib) |
| | c.1924A > G/p.K642E[1] | 1 (0.4 %) | | |
| | c.1727T > C/p.L576P[1] | 2 (0.7 %) | | |
| KRAS | c.35G > A/p.G12D[1] | 1 (0.4 %) | 4 (1.5 %) | AZD6244 (Selumetinib), U0126-EtOH, PD98059, AS703026, MEK pathway inhibitors |
| | c.181C > A/p.Q61K[1] | 1 (0.4 %) | | |
| | c.182A > T/p.Q61L[1] | 1 (0.4 %) | | |
| | c.183A > C/p.Q61H[1] | 1 (0.4 %) | | |
| PDGFR-α | c.1977C > A/p.N659K[1] | 1 (0.4 %) | 3 (1.1 %) | IMC-3G3[a], MEDI 575[a], Pazopanib Hydrochloride (GW786034, VOTRIENT), Sunitinib Malate (Sutent), PKC412 (Midostaurin o CGP 41251), Gleevec (Imatinib mesylate, STI-571), Axitinib (AG-013736), ZD6474 (Vandetanib), Sorafenib Tosylate (Nexavar, or Bay 43-9006), Dasatinib (BMS-354825, Sprycel, o BMS354825), Nilotinib (AMN107, MP470, Masitinib mesylate (AB1010), Tandutinib, Vargatef (BIBF1120), AG18, AG-213 (Tyrphostin AG 213), TSU-68 (SU6668), AP24534 (Ponatinib), Imatinib mesylate (Gleevec) |
| | c.2470G > C/p.V824L[1] | 2 (0.7 %) | | |

**Table 2** continued

| Gene symbol | Genomic and codon positions, base and amino acid substitutions* | Number (%) of cases with mutations at allele level | Number (%) of cases with mutation at gene level | Drugs that inhibit gene function or block the involved signaling pathway |
|---|---|---|---|---|
| **PHLPP2** | **c.3047T > C/p.L1016S**[2] | 36 (13.5 %) | 36 (**13.5 %**) | Reduced PHLPP2 activity has similar functional consequences as PTEN loss and these tumor cells are more sensitive to AKT inhibition |
| PIK3CA | c.328G > A/p.E110K[1] | 3 (1.1 %) | 43 (**16.1 %**) | XL147, GDC0941, BKM120, PX866, CAL101, BYL719, GSK1059615, XL765, BEZ235, SF1126, BGT226, GSK2126458, PIK90, NVP-BBD130, LY294002, NVP-BAG956, NVP-BBD130, GDC-0941 bismesylate, PIK 75 hydrochloride, PI-103, ZSTK474, AS 605240, NVP-BEZ235, Everolimus (RAD001), Temsirolimus (CCI-779), Rapamycin (Sirolimus) |
| | c.1357G > A/p.E453K[1] | 1 (0.4 %) | | |
| | c.1624G > A/p.E542K[1] | 4 (1.5 %) | | |
| | c.1633G > A/p.E545K[1] | 4 (1.5 %) | | |
| | c.1634A > C/p.E545A[1] | 4 (1.5 %) | | |
| | c.3140A > G/p.H1047R[1] | 14 (5.1 %) | | |
| | c.3073A > T/p.T1025S[1] or c.3073A > G/p.T1025A[1] | 13 (4.8 %) | | |
| PIK3R1 | c.1678G > T/p.D560Y[1] | 1 (0.4 %) | 1 (**0.4 %**) | PI3K-mTOR pathway inhibitors e.g. NVP-BAG956, PI-828, LY294002, Everolimus (RAD001), Temsirolimus (CCI-779), Rapamycin (Sirolimus) |
| RET | c.2753T > C/p.M918T[1] | 1 (0.4 %) | 1 (**0.4 %**) | ZD6474 (Vandetanib), XL184, Sorafenib Tosylate (Nexavar, or Bay 43-9006), Sunitinib Malate (Sutent), PHA-739358 (Danusertib), BAY 73-4506 (Regorafenib), AMG-706 (Motesanib Diphosphate) |
| TNK2/ACK1 | c.1036G > A/p.E346K[1] | 1 (0.4 %) | 1 (**0.4 %**) | EGFR/AKT pathway inhibitors |

* c = genomic position; > = nucleotide changes (nucleotides: *A* adenine, *C* cytosine, *G* guanine, and *T* thymine); *p* protein, codon position and aminoacids: *A* alanine, *R* arginine, *N* asparagine, *D* aspartic acid, *C* cysteine, *E* glutamic acid, *Q* glutamine, *G* glycine, *H* histidine, *I* isoleucine, *L* leucine, *K* lysine, *M* methionine, *F* phenylalanine, *P* proline, *S* serine, *T* threonine, *W* tryptophan, *Y* tyrosine, and *V* valine. (*e.g.*: AKT "c.49G > A/p.E17 K" indicates nucleotide change at location 49 from guanine to adenine with a consequential aminoacid change from glutamic acid to lysine at codon 17)

*AKT1* v-akt murine thymoma viral oncogene homolog 1, *BRAF* v-raf murine sarcoma viral oncogene homolog B1, *CTNNB1* catenin (cadherin-associated protein), beta 1, *EGFR* epidermal growth factor receptor, *FBXW7* F-box and WD repeat domain containing 7, *HIF1-α* hypoxia inducible factor 1, alpha subunit, *IDH2* isocitrate dehydrogenase 2 (NADP+), mitochondrial, *KIT* v-Kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog, *KRAS* v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog, *PDGFR-α* platelet-derived growth factor receptor, *PIK3CA* phosphoinositide-3-kinase, catalytic, alpha polypeptide, *PIK3R1* phosphoinositide-3-kinase, regulatory subunit 1 (alpha), *RET* ret proto-oncogene, *TNK2/ACK1* tyrosine kinase, non-receptor, 2

[1]  known functionally activating mutation

[2]  known functionally inactivating mutations also printed *bold*

[3]  unknown functional relevance

[a]  Monoclonal antibody

Synonymous names are in parenthesis

Eighty-one of the 108 samples (30 %) had one mutation, 23 samples (9 %) had two, and four samples (1 %) had ≥ three mutations (Supplementary Fig. 1). The most frequently mutated or activated genes were *PIK3CA* (n = 43), *PHLPP2* (n = 36), and *FBXW7* (n = 21) (Table 2). Figure 1 shows a heat map of all gene-level nucleic acid variations annotated by routine clinical variables and samples that showed concurrent mutations.

When gene-level mutation frequencies were compared across the three clinical subtypes of breast cancer, *PIK3CA* mutations were significantly more common in ER+ cancers compared to TNBC (19 *vs.* 8 %, *p = 0.001*) (Table 3). *PIK3CA* were also commonly mutated in HER2+ cancers (28 %). In TNBC, *FBXW7* mutations were significantly more frequent compared to ER+ cancers (13 *vs.* 5 %, *p = 0.037*) (Table 3).
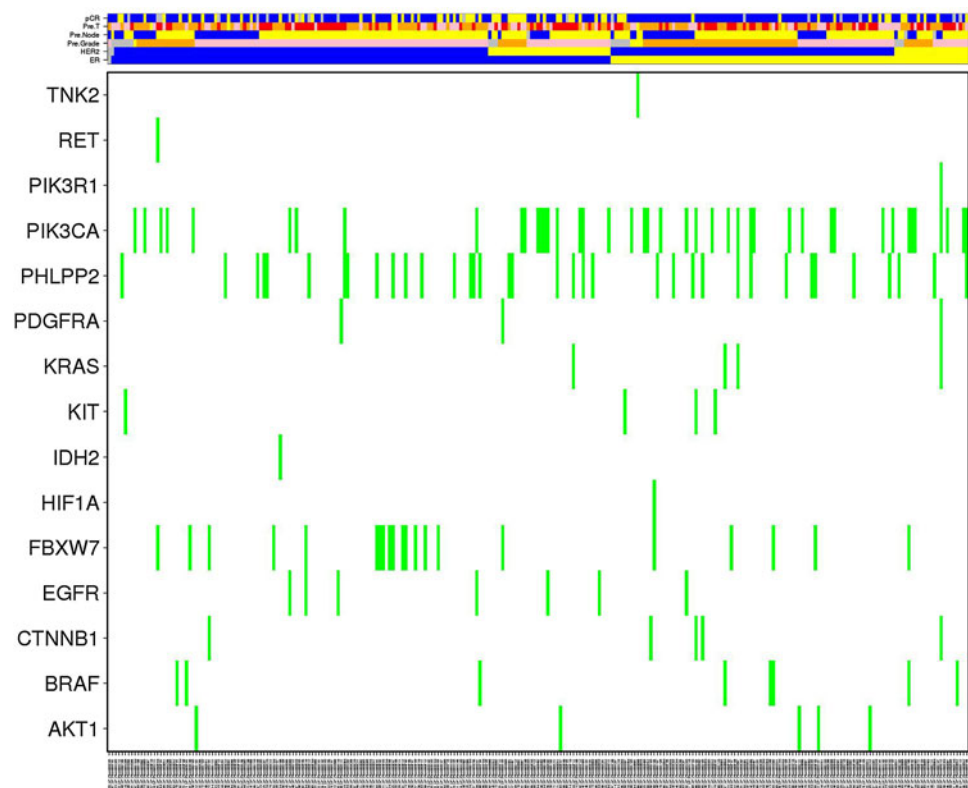
When gene-level mutations were collapsed into three mechanistic groups based on the known function of the genes, 32 % of cases had a mutation or activation in genes (*AKT1, PHLPP2, PIK3CA,* and *PIK3R1*) that could potentially lead to activation of the PI3K pathway, 6 % had activating mutations in at least one of the following RTKs (*EGFR, KIT, RET,* and *PDGFR-α*), and 15 % had a mutation in at least one gene involved in the regulation of cell cycle and metabolism (*BRAF, CTNNB1, FBXW7, HIF1-α, IDH2, KRAS,* and *TNK2*) (Fig. 2). Mutations in the PI3K pathway were significantly more common in ER+

and HER2+ cancers compared to TNBC (*p = 0.032*, Table 3). Multivariate logistic regression analysis confirmed significant positive associations between PI3K pathway mutations and ER (odds ratio [OR]:2.18, *p = 0.01*) and HER2 status (OR: 2.22, *p = 0.02*), respectively (Table 4). In multivariate analysis, mutations in RTKs had a positive trend to be associated with HER2 status (OR: 3.24, *p = 0.09*). Mutations in the cell cycle/metabolism genes showed significant inverse association with tumor size (OR: 0.57, *p = 0.01*) and a positive association with nodal status (OR: 2.61, *p = 0.05*).

### Association between mutation status and clinical outcome

None of the individual genes alone had significant association with chemotherapy response or survival. However, at pathway level, mutation in the PI3K pathway was associated with higher probability of pCR to preoperative chemotherapy (OR: 2.04, *p = 0.04*), in multivariate analysis (Table 4). Multivariate analysis for subgroups by ER-status could not be performed because of the small cohort sizes in each subset. The logrank *p*-values of the comparisons between the K–M plots of the mutant versus wild type for each pathway and each mutated gene are shown in Supplementary Fig. 2–4. None of the survival curve comparisons were significantly different.



**Fig. 1** Heat-map of gene level mutations across all samples (n = 267). The *x-axis* represent tumor samples, the *y-axis* shows gene names, white indicates wild type and green mutated gene. The samples were ordered first by ER status, then by HER2 status, tumor grade (PreGrade), tumor node (PreNode), tumor size status (PreT), and pathological complete response (pCR) as shown on the rug plot on the top. Color scheme in the rug plot: *Gray* = missing values, for ER, HER2, PreNode and pCR status *blue* = negative and *yellow* = positive, for PreGrade and PreT levels *yellow* = 1, *orange* = 2, *pink* = 3, *red* = 4, where the numbers correspond to histologic grade and tumor size by TNM staging

| | ER +/HER2− (n = 88) (%) | HER2+ (n = 61) (%) | TN (n = 116) (%) | FET p-value |
|---|---|---|---|---|
| PI3K pathway | 29 (33) | 24 (39.3) | 25 (21.6) | **0.03** |
| RTKs pathway | 4 (4.5) | 4 (6.6) | 7 (6) | 0.838 |
| CCM pathway | 11 (12.5) | 5 (8.2) | 19 (16.4) | 0.32 |
| AKT1 | 3 (3.4) | 1 (1.6) | 1 (0.9) | 0.449 |
| BRAF | 3 (3.4) | 2 (3.3) | 3 (2.6) | 1 |
| CTNNB1 | 3 (3.4) | 1 (1.6) | 1 (0.9) | 0.449 |
| EGFR | 1 (1.1) | 2 (3.3) | 4 (3.4) | 0.614 |
| FBXW7 | 4 (4.5) | 2 (3.3) | 15 (12.9) | **0.037** |
| HIF1A | 1 (1.1) | 0 (0) | 0 (0) | 0.562 |
| IDH2 | 0 (0) | 0 (0) | 1 (0.9) | 1 |
| KIT | 3 (3.4) | 0 (0) | 1 (0.9) | 0.279 |
| KRAS | 2 (2.3) | 2 (3.3) | 0 (0) | 0.119 |
| PDGFRA | 0 (0) | 2 (3.3) | 1 (0.9) | 0.169 |
| PHLPP2 | 11 (12.5) | 9 (14.8) | 16 (13.8) | 0.922 |
| PIK3CA | 17 (19.3) | 17 (27.9) | 9 (7.8) | **0.001** |
| PIK3R1 | 0 (0) | 1 (1.6) | 0 (0) | 0.23 |
| RET | 0 (0) | 0 (0) | 1 (0.9) | 1 |
| TNK2 | 1 (1.1) | 0 (0) | 0 (0) | 0.562 |

**Table 3** Individual gene and signalling pathway mutation frequencies in the three main different breast cancer subtypes

*FET* Fisher's Exact Test, *p* values were unadjusted for multiple comparisons, significant values are in *bold*. *PI3K* Phosphatidylinositol, *RTK* Receptor tyrosine kinase, *CCM* Cell cycle/Metabolic

## Validation of Sequenom findings with different methods

We performed independent technical validation of all mutated alleles for 15 genes. *PIK3CA* mutation status was available for a subset of this cohort from a previous study, and these results were also included in the validation [16]. Mutation results that did not reach a concordance of at least 95 % between different techniques were discarded and not reported on. The Sequenom methodology used in these experiments were subsequently transferred to a CLIA-certified laboratory environment and are now performed as routine clinical assays.

## Discussion

Overall, our results indicate that breast cancers harbor several individually rare but potentially targetable mutations. Activating mutations in the *PIK3CA* gene were the most common but we also observed activating mutations of *AKT1*, *BRAF*, *EGFR*, *KIT*, *KRAS*, and *PDGFR-α*. The functional importance of these mutations in human breast cancer remains unknown; however, each of these genes can influence cancer growth in at least some experimental model systems and they each can be targeted with approved or investigational drugs (Table 2). The ultimate tests to determine whether any of these mutated genes represent therapeutic targets in breast cancer are a series of clinical trials. Our data demonstrated that it is feasible to

determine the mutation status for this panel of genes in tissues from fine needle aspirations and several academic institutions have initiated projects to set up mutation profiling of solid tumors in routine pathology laboratories (http://www.mdanderson.org/education-and-research/resources-for-professionals/scientific-resources/core-facilities-and-services/molecular-diagnostics-lab/services/index.html). These developments provide a logistical framework to conduct such studies in the near future.

A potential limitation of this study is that we did not analyze normal samples matching the cancer specimens; however, these mutations all have been previously described in at least some cancer specimens as somatic events. We also included in our mutation panel functionally relevant SNPs such as *PHLPP2* (PH Domain and Leucine Rich Repeat Protein Phosphatase 2). We observed a variant nucleotide in 13.5 % of cancers for this gene (c.3047T > C/p.L1016S). This variant is known to inactivate the phosphatase activity of *PHLPP2*, which dephosphorylates AKT-1,-2,-3 and protein kinase C and there is laboratory evidence to indicate that this leads to sustained AKT signaling in breast cancer cell lines [17]. It is therefore plausible that this germ-line variant may also influence the behavior of breast cancer in patients and may even represent a strategy to select patients for therapy with AKT pathway inhibitors.

Several other clinically testable therapeutic hypotheses can be formulated based on the mutations that we detected. We noted inactivating mutations in *FBXW7* (*F*-box and WD Repeat Domain Containing-7) in 8 % of cases, and it was
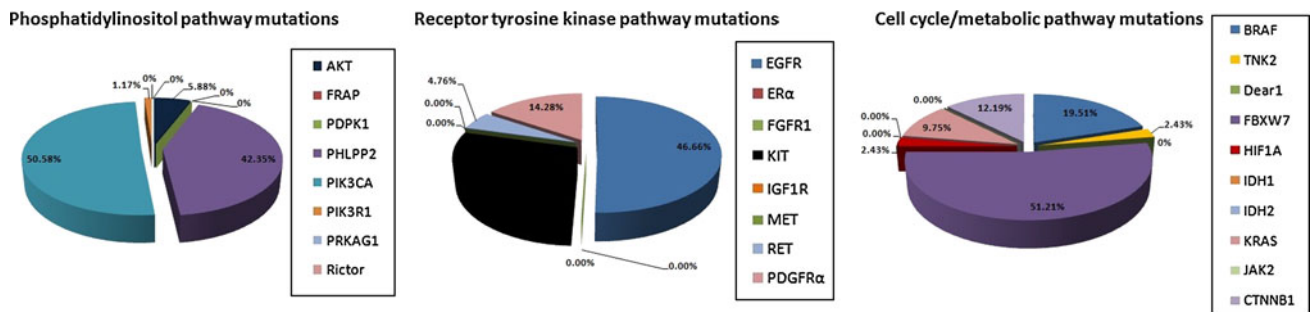
**Fig. 2** Gene level mutations grouped into 3 functional categories. The percentages of individual gene level mutations that contribute to the three distinct functional categories are shown

**Table 4** Association between clinical variables and mutation status at pathway level

| | Fitted beta | Odds ratio | OR. LCI | OR. UCI | *p* value |
|---|---|---|---|---|---|
| Multivariate logistic regression analysis results for PI3K pathway | | | | | |
| ER | 0.78 | 2.18 | 1.16 | 4.07 | 0.0149* |
| HER2 | 0.80 | 2.22 | 1.09 | 4.52 | 0.0285* |
| Grade 2/3 | −0.34 | 0.71 | 0.11 | 4.62 | 0.7201 |
| *T* | 0.03 | 1.03 | 0.73 | 1.45 | 0.8775 |
| *N* | −0.61 | 0.54 | 0.28 | 1.05 | 0.0703 |
| pCR | 0.71 | 2.04 | 1.01 | 4.11 | 0.0461* |
| Multivariate logistic regression analysis results for receptor tyrosine kinase pathway | | | | | |
| ER | −0.30 | 0.74 | 0.20 | 2.78 | 0.6534 |
| HER2 | 1.18 | 3.24 | 0.81 | 13.04 | 0.0975 |
| Grade 2/3 | 17.35 | 34131509.05 | 0.00 | Inf | 0.9982 |
| *T* | 0.18 | 1.20 | 0.56 | 2.57 | 0.6373 |
| *N* | −0.18 | 0.84 | 0.19 | 3.70 | 0.8161 |
| pCR | −18.12 | 0.00 | 0.00 | Inf | 0.9936 |
| Multivariate logistic regression analysis results for cell cycle/metabolic pathway | | | | | |
| ER | −0.05 | 0.95 | 0.43 | 2.10 | 0.9019 |
| HER2 | −0.22 | 0.80 | 0.28 | 2.33 | 0.6852 |
| Grade 2/3 | −0.10 | 0.91 | 0.08 | 10.03 | 0.9379 |
| *T* | −0.56 | 0.57 | 0.36 | 0.89 | 0.0143* |
| *N* | 0.96 | 2.61 | 0.99 | 6.86 | 0.0520* |
| pCR | −0.44 | 0.65 | 0.25 | 1.69 | 0.3724 |

*OR* odds ratio for having mutations in pathway versus no-mutations. *L* lower and *U* upper limits of the 95% confidence intervals of the OR are also shown

* Indicate significant associations

significantly more common in TNBC. This gene codes for a substrate-recognition factor that targets ubiquitin ligase to c-Myc, cyclin E, c-Jun, TGF-$\beta$, Notch, and mTOR for degradation [18]. Tumor cell lines harboring deletions or inactivating mutations in *FBXW7* have high mTOR levels and are particularly sensitive to mTOR inhibitors [19].

In summary, this analysis shows that about 40 % of breast cancers contain mutations in common signaling pathways that can be targeted with currently clinically available drugs. Different breast cancer subtypes harbor different types of mutations, but any given mutation affects only a small subset of cases. Simultaneous testing of many different mutations in a single needle biopsy is feasible and allows the design of innovative prospective clinical trials

that could test the functional importance of these mutations and the clinical value of mutation-based patient selection strategies for biologically targeted drugs.

**Author Contributions** LP and LS designed the study, oversaw all of its conduct, and finalized the manuscript. LS and YQ carried out the analysis of data and performed statistical analysis, LS drafted the manuscript; BW and KS-H performed molecular studies on specimens. DJH, FAH, JO'Sm GH, JP, TV, HG, VV, GNH, AMG accrued

patients to the study, made biopsy materials available for molecular analyis and provided clinical input. EJY maintained the clinical data base. WFS obtained many of the cancer biopsies for this study. ADL participated in the design and provided intellectual input into interpretation of results. GB provided intellectual input into interpretation of results. All authors have read and approved the final version.

**Conflict of interest** The authors declare no conflicting or competing commercial interests.

# References

1. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, Buck G, Chen L, Beare D, Latimer C, Widaa S, Hinton J, Fahey C, Fu B, Swamy S, Dalgliesh GL, Teh BT, Deloukas P, Yang F, Campbell PJ, Futreal PA, Stratton MR (2010) Signatures of mutation and selection in the cancer genome. Nature 463:893–898

2. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning Z, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L, Ross MT, Campbell PJ, Bentley DR, Futreal PA, Stratton MR (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463:191–196

3. Teschendorff AE, Caldas C (2009) The breast cancer somatic 'muta-ome': tackling the complexity. Breast Cancer Res 11:301

4. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR (2007) Patterns of somatic mutation in human cancer genomes. Nature 446:153–158

5. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B (2007) The genomic landscapes of human breast and colorectal cancers. Science 318:1108–1113

6. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 461:809–813

7. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, Wang M, Feng W, Zander T, MacConaill L, Lee JC, Nicoletti R, Hatton C, Goyette M, Girard L, Majmudar K, Ziaugra L, Wong KK, Gabriel S, Beroukhim R, Peyton M, Barretina J, Dutt A, Emery C, Greulich H, Shah K, Sasaki H, Gazdar A, Minna J, Armstrong SA, Mellinghoff IK, Hodi FS, Dranoff G, Mischel PS, Cloughesy TF, Nelson SF, Liau LM, Mertz K, Rubin MA, Moch H, Loda M, Catalona W, Fletcher J, Signoretti S, Kaye F, Anderson KC, Demetri GD, Dummer R, Wagner S, Herlyn M, Sellers WR, Meyerson M, Garraway LA (2007) High-throughput oncogene mutation profiling in human cancer. Nat Genet 39:347–351

8. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerød A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Børresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 462:1005–1010

9. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O'Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464:999–1005

10. MacConaill LE, Campbell CD, Kehoe SM, Bass AJ, Hatton C, Niu L, Davis M, Yao K, Hanna M, Mondal C, Luongo L, Emery CM, Baker AC, Philips J, Goff DJ, Fiorentino M, Rubin MA, Polyak K, Chan J, Wang Y, Fletcher JA, Santagata S, Corso G, Roviello F, Shivdasani R, Kieran MW, Ligon KL, Stiles CD, Hahn WC, Meyerson ML, Garraway LA (2009) Profiling critical cancer gene mutations in clinical tumor samples. PLoS ONE 4:e7887

11. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L, Wang W, Eigenbrot C, Kaminker JS, Eberhard DA, Waring P, Schuster SC, Modrusan Z, Zhang Z, Stokoe D, de Sauvage FJ, Faham M, Seshagiri S (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 466:869–873

12. Symmans WF, Ayers M, Clark EA, Stec J, Hess KR, Sneige N, Buchholz TA, Krishnamurthy S, Ibrahim NK, Buzdar AU, Theriault RL, Rosales MF, Thomas ES, Gwyn KM, Green MC, Syed AR, Hortobagyi GN, Pusztai L (2003) Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. Cancer 97:2960–2971

13. Jurinke C, Oeth P, van den Boom D (2004) MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. Mol Biotechnol 26:147–163

14. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M, Qi Y, Barajas-Figueroa LJ, Souchon E, Coutant C, Doimi FD, Ibrahim NK, Gong Y, Hortobagyi GN, Hess KR, Symmans WF, Pusztai L (2010) Evaluation of a 30-gene paclitaxel, fluorouracil,

doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. Clin Cancer Res 16:5351–5361

15. Iwamoto T, Bianchini G, Booser D, Qi Y, Coutant C, Shiang C, Santarpia L, Matsuoka J, Hortobagyi GN, Symmans WF, Holmes FA, O'Shaughnessy J, Hellerstedt B, Pippen J, Andre F, Simon R, Pusztai L (2011) Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. J Natl Cancer Inst 103:264–272

16. Liedtke C, Cardone L, Tordai A, Yan K, Gomez HL, Figureoa LJ, Hubbard RE, Valero V, Souchon EA, Symmans WF, Hortobagyi GN, Bardelli A, Pusztai L (2008) PIK3CA-activating mutations and chemotherapy sensitivity in stage II–III breast cancer. Breast Cancer Res 10:R27

17. Brognard J, Niederst M, Reyes G, Warfel N, Newton AC (2009) Common polymorphism in the phosphatase PHLPP2 results in reduced regulation of Akt and protein kinase C. J Biol Chem 284:15215–15223

18. Welcker M, Clurman BE (2007) FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. Nat Rev Cancer 8:83–93

19. Mao JH, Kim IJ, Wu D, Climent J, Kang HC, DelRosario R, Balmain A (2008) FBXW7 targets mTOR for degradation and cooperates with PTEN in tumor suppression. Science 321:1499–1502