

# Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival

Victor J. Weigman · Hann-Hsiang Chao · Andrey A. Shabalin · Xiaping He · Joel S. Parker · Silje H. Nordgard · Tatyana Grushko · Dezheng Huo · Chika Nwachukwu · Andrew Nobel · Vessela N. Kristensen · Anne-Lise Børresen-Dale · Olufunmilayo I. Olopade · Charles M. Perou

Received: 2 October 2011 / Accepted: 4 October 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** Breast cancer is a heterogeneous disease with known expression-defined tumor subtypes. DNA copy number studies have suggested that tumors within gene expression subtypes share similar DNA Copy number aberrations (CNA) and that CNA can be used to further subdivide expression classes. To gain further insights into the etiologies of the intrinsic subtypes, we classified tumors

according to gene expression subtype and next identified subtype-associated CNA using a novel method called SWITCHdna, using a training set of 180 tumors and a validation set of 359 tumors. Fisher's exact tests, Chi-square approximations, and Wilcoxon rank-sum tests were performed to evaluate differences in CNA by subtype. To assess the functional significance of loss of a specific chromosomal region, individual genes were knocked down by shRNA and drug sensitivity, and DNA repair foci assays performed. Most tumor subtypes exhibited specific CNA. The Basal-like subtype was the most distinct with common losses of the regions containing *RBI*, *BRCA1*, *INPP4B*, and the greatest

Victor J. Weigman and Hann-Hsiang Chao contributed equally to this study.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10549-011-1846-y) contains supplementary material, which is available to authorized users.

V. J. Weigman  
Bioinformatics and Computational Biology Program,  
University of North Carolina, Chapel Hill, NC 27599, USA

V. J. Weigman · X. He · J. S. Parker · C. M. Perou (✉)  
Lineberger Comprehensive Cancer Center, University of North Carolina, 450 West Drive, CB7295, Chapel Hill, NC 27599, USA  
e-mail: cperou@med.unc.edu

V. J. Weigman  
Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

H.-H. Chao · X. He · J. S. Parker · C. M. Perou  
Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

A. A. Shabalin · A. Nobel  
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

A. A. Shabalin · A. Nobel  
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

S. H. Nordgard · V. N. Kristensen · A.-L. Børresen-Dale  
Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Norway

T. Grushko · D. Huo · C. Nwachukwu · O. I. Olopade  
Center for Clinical Cancer Genetics and Global Health, University of Chicago Medical Center, MC 2115, Chicago, IL 60615, USA

V. N. Kristensen  
Department of Clinical Molecular Biology (EpiGen), Akerhus University Hospital, University of Oslo, Oslo, Norway

V. N. Kristensen · A.-L. Børresen-Dale  
Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

C. M. Perou  
The Carolina Genome Sciences Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

C. M. Perou  
Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

overall genomic instability. One Basal-like subtype-associated CNA was loss of 5q11–35, which contains at least three genes important for BRCA1-dependent DNA repair (*RAD17*, *RAD50*, and *RAP80*); these genes were predominantly lost as a pair, or all three simultaneously. Loss of two or three of these genes was associated with significantly increased genomic instability and poor patient survival. RNAi knockdown of *RAD17*, or *RAD17/RAD50*, in immortalized human mammary epithelial cell lines caused increased sensitivity to a PARP inhibitor and carboplatin, and inhibited BRCA1 foci formation in response to DNA damage. These data suggest a possible genetic cause for genomic instability in Basal-like breast cancers and a biological rationale for the use of DNA repair inhibitor related therapeutics in this breast cancer subtype.

**Keywords** Basal-like breast cancer · Genome instability · BRCA1 pathway · Copy number aberration · Molecular subtypes · Array CGH

### Abbreviations

CNA	Copy number aberrations
MTT	3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide
aCGH	Array comparative genomic hybridization
HME-CC	hTERT-immortalized human mammary epithelial cell line
ME16C	hTERT-immortalized human mammary epithelial cell line
UMD	UNC microarray database
UNC	University of North Carolina
USA	Samples
NW	Norway samples
DWD	Distance weighted discrimination
FWER	Familywise error rate
pCR	Pathologic complete response
T/FAC	Taxane, fluorouracil, anthracycline, cyclophosphamide

### Introduction

Previous gene expression profiling studies of human breast tumors have shaped our understanding that breast cancer is not one disease, but is in fact many biologically separate diseases. A classification of tumors by expression profiling into five distinct groups (Luminal A, Luminal B, HER2-enriched, Basal-like, and Claudin-low subtypes) has added prognostic and predictive value to the existing repertoire of biomarkers for breast cancer [1–6]. For many cancers, improper maintenance of genome stability is a major cause of tumorigenesis and thus, the characterization of the tumor genomic DNA landscape is an important avenue of

investigation [7]. Array comparative genome hybridization (aCGH) studies of tumor copy number states have demonstrated that tumors with similar gene expression subtypes may also share similar DNA copy number aberrations (CNA) [8–12] and that CNA can be used to further subdivide expression classes [12]. In breast cancers, genomic instability-driven tumorigenesis is most prevalent in the Basal-like subtype (also referred to as triple-negative breast cancers), where the majority of tumors exhibit many CNA [9–13]. Identifying the genes that contribute to this instability phenotype would be useful not only from a biological perspective, but also possibly as a clinical predictor of therapeutic response.

### Methods

A detailed description of all methods is provided in the “[Supplemental Methods](#)” section, while here we provide an abbreviated methods section for the major new approaches.

#### Breast cancer patient datasets

For the genomic studies, three patient datasets were used, each containing gene expression and DNA copy number microarray data. We combined two sets into a single training set ( $n = 180$  with expression and copy number) so that we could have increased statistical power to detect subtype-specific CNA. The combined training set included breast tumors from the United States (“UNC”) ( $n = 77$ ) and tumors from Norway (“NW”) ( $n = 103$ ). The third data set (“Jonsson”) was used as a validation/testing set ( $n = 359$ ) [14]. All samples were collected using IRB-approved protocols. Data is available from Gene Expression Omnibus series GSE10893. Sample information including clinical data, subtype, source, GEO Sample ID, and overlap with copy number information can be found in Supplemental Table 1.

#### Assessment of tumor genomic DNA copy number changes

77 UNC and 103 NW samples had normal and tumor DNA samples each assayed using the Infinium Human-1 109K BeadChip (Illumina, San Diego, CA, USA). Sample information is provided in Supplemental Table 1 and LogR (A+B signal) values can be found on GEO series GSE10893, platform GPL8139. To determine regions of copy number aberration (CNA), we developed a new analysis method that is a modification of the SupWald method [15, 16]; we created an R suite of functions called “SWITCHdna”, which can identify breakpoints in aCGH data. SWITCHdna detects transition points that maximize

the  $F$  statistic and have regions on either side of the breakpoint that are larger than the user-defined range. Following detection of the transition points, a segment average value and corresponding  $z$ -score are determined, along with the number of observations used. The end results are the identification of segments of CNA, along with a quantitative value for that copy number change (i.e., loss or gain).

A significance filter is applied to the raw SWITCHdna-identified segments in order to reduce noise and increase the probability of identifying biologically relevant regions. All subsequent plots and tables were produced after applying this significance filter to our data. SWITCHdna is provided as a source script in R [17] and available for download at: <https://genome.unc.edu/pubsup/SWITCHdna/>.

### Determining subtype-specific CNAs

Using the `cnaGENE` function of SWITCHdna, the segment output file was converted into an indicator matrix, where for each sample, each gene's copy state was represented as  $-1 = \text{loss}$ ,  $0 = \text{no change}$ ,  $1 = \text{gain}$ . For each subtype, the counts of gains and losses were compared versus all other samples in order to identify subtype-specific CNAs. A Fisher's exact test was performed on the subtype versus rest counts for each gene. The resulting  $P$  values were adjusted by the Benjamini-Hochberg method [18] to correct for multiple-hypothesis testing and genes with  $P$  values  $< 0.05$  were then gathered for each subtype. Regions within the cytobands of localized CNA were determined by the significant genes found within each cytoband (Supplemental Table 2).

### Supplemental methods

Numerous additional methods, and more detail on SWITCHdna is provided in the Supplemental Methods section. These methods include details on the cell lines used, RNAi knockdown experiments, and other cell biology type experiments performed here.

## Results

### Identifying subtype-specific regions of copy number aberration

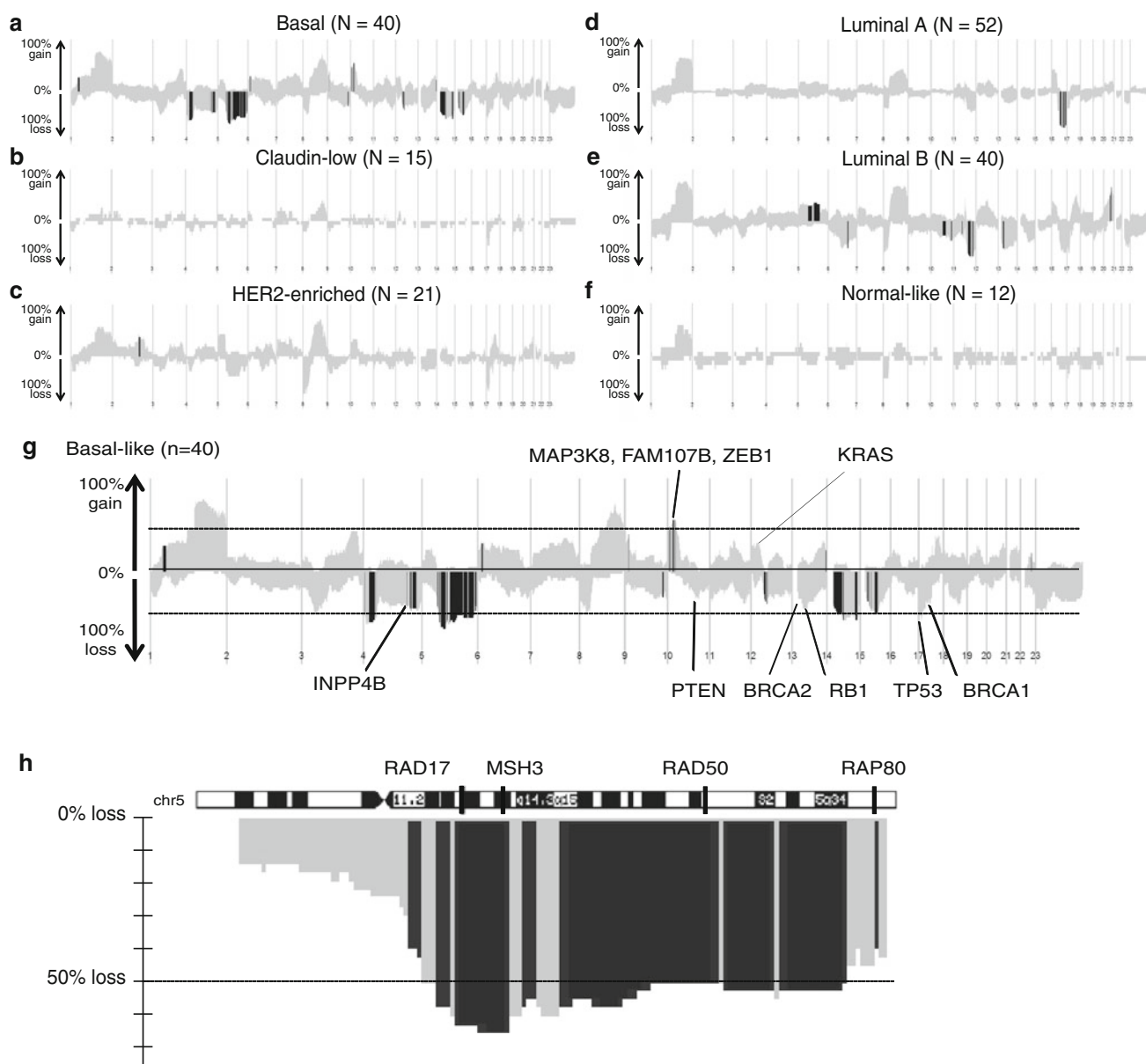
To identify CNA that might be causative of Basal-like breast cancers, we assembled a dataset of 180 tumors with Agilent gene expression microarrays and Illumina 109,000 SNP marker DNA copy number microarrays (UNC-NW). We classified each tumor into one of five previously defined expression subtypes using the published intrinsic subtypes (i.e., PAM50) and Claudin-low subtype predictors

[5, 6]. To identify regions of copy number gain/loss, we developed a new segmenting method called "SWITCHdna" (Sup Wald Identification of copy CHanges in dna). Specifics of the SWITCHdna method can be found in the "Supplemental Methods" and at <https://genome.unc.edu/pubsup/SWITCHdna/>.

SWITCHdna-identified regions/segments of copy number gains and losses in each tumor, which were then aggregated based on subtype to look at the frequency of each copy number event in each subtype and identify regions specific to each subtype (Fig. 1; Supplemental Table 2). A heat map display of the copy number data is provided in Supplemental Fig. 1. A number of new findings were observed including the first aCGH characterization of the Claudin-low subtype (Fig. 1b). Despite its high grade and similarity to Basal-like tumors [5, 6], Claudin-low tumors showed few copy number changes, which may correspond to the previously described ER-negative and copy number neutral tumor subtype reported in Chin et al. [19]. In addition, human Claudin-low cell lines, which are often called "Basal B" lines, also have a similar flat copy number profile of showing very few chromosomal abnormalities [20].

We next searched for CNA occurring specifically within each subtype (Fig. 1a–f, black shading). The Basal-like subtype had the most subtype-specific events (Fig. 1a, g) including the previously described amplicon at 10p containing *MAP3K8*, *ZEB1*, and *FAM107B* [13, 21, 22], 16q loss [23], deletion of 5q11–35 [10], and deletion of 4q. This last region contains *INPP4B*, which has recently been identified as a potential tumor suppressor involved in the inhibition of PI3K signaling [24] and that is selectively lost in Basal-like/Triple-negative breast cancers [25].

Basal-like tumors have previously been observed to have copy number loss and/or low expression of genes involved in BRCA1 DNA damage repair [26], and we noted that loss of 5q11–5q35 would delete several genes involved in BRCA1-dependent DNA repair including *RAD17*, *RAD50* [27], and *RAP80* (Fig. 1h). Closer examination of the pattern of loss of these genes revealed that each gene was rarely lost as an individual event, but predominantly lost as a pair or triplet (Table 1a). These doublet or triplet losses occurred at the highest rates in the Basal-like subtype, but also occurred less frequently in the HER2-enriched subtype. These paired or triplet losses were not simply due to loss of the entire chromosomal arm as  $>65\%$  of the analyzed tumors did not show a loss pattern indicative of such an event and several samples had intervening regions of normal copy number. Loss of 5q11–35 was also found to statistically co-occur with CNA of other regions including 10p amplification ( $\sim 50\%$ ), *INPP4B*/4q31.21 loss ( $\sim 40\%$ ), *PTEN*/10q23.31 loss ( $\sim 40\%$ ), *BRCA1*/17q21 loss ( $\sim 50\%$ ), and most frequently



**Fig. 1** Copy number frequency plots from SWITCHdna show regions of aberrations shared by members of the same subtype. *Gray* shading indicates regions of change with the y-axis representing frequency of aberration at each site within each subtype. Regions in *black* were statistically associated with a particular subtype and remained significant after Benjamini-Hochberg correction. Regions below the center (negative values) represent losses, and areas above the center (positive values) indicate gains. **a** Basal-like, **b** Claudin-low, **c** HER2-enriched, **d** Luminal A, **e** Luminal B, and **f** Normal-like. **g** Expanded view of the Basal-like copy number landscape. *INPP4B*,

*MAP3K8*, *FAM107B*, and *ZEB1*, each in Basal-like specific regions of CNA, are marked. *BRCA1*, *BRCA2*, *PTEN*, *RB1*, and *TP53*, are genes/regions that were frequently, but not specifically, lost in the Basal-like subtype, and *KRAS*, which is frequently but not specifically gained in the Basal-like subtype, are also noted. The *dashed horizontal lines* indicate 50% gain or loss. **h** Enlarged view of the Basal-like chromosome 5q region showing the location of *RAD17*, *MSH3*, *RAD50*, and *RAP80*. Loss frequency is indicated on the y-axis and the level of 50% loss is highlighted by the *horizontal line*

loss of *RB1*/13q14.2 (~80%) (Table 1e), which are genes/regions that have all previously shown to be associated with Basal-like breast cancers.

In order to validate these subtype-specific findings observed in the UNC+NW dataset, we classified the samples in Jonsson et al. [14] according to PAM50 and Claudin-low

subtype predictors and performed similar supervised analyses using their BAC-based DNA copy number data; very similar associations between CNA and subtypes were observed (Table 2). Jonsson et al. identified six unique tumor subtypes based upon CNA landscapes, which we determined were highly correlated with our expression-defined intrinsic

**Table 1** Frequency of copy number alterations data for the UNC-Norway combined dataset for selected (a) deletions, (b) amplifications, (c) average number of changes, (d) % Tumor Cellularity, and (e) co-occurrences

	UNC-NW												All ( <i>n</i> = 180)		
	Basal ( <i>n</i> = 40)		Claudin ( <i>n</i> = 15)		Her2 ( <i>n</i> = 21)		LumA ( <i>n</i> = 52)		LumB ( <i>n</i> = 40)		Normal-like ( <i>n</i> = 12)		<i>P</i> -Value <sup>a</sup>	No.	%
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%			
<b>(a) Deletions</b>															
No 5q Genes Lost	13	32.5	11	73.3	11	52.4	42	80.8	32	80.0	10	83.3	<0.001	119	66.1
5q13.2 (RAD17)	5	12.5	1	6.7	1	4.8	4	7.7	2	5.0	1	8.3	<0.001	14	7.8
5q31.1 (RAD50)	1	2.5	0	0.0	1	4.8	0	0.0	2	5.0	0	0.0	<0.001	4	2.2
5q35.2 (RAP80)	1	2.5	1	6.7	0	0.0	2	3.8	2	5.0	0	0.0	0.001	6	3.3
RAD17/RAD50	5	12.5	2	13.3	3	14.3	2	3.8	0	0.0	0	0.0	<0.001	12	6.7
RAD17/RAP80	1	2.5	0	0.0	0	0.0	0	0.0	1	2.5	1	8.3	<0.001	3	1.7
RAD50/RAP80	0	0.0	0	0.0	0	0.0	1	1.9	1	2.5	0	0.0	<0.001	2	1.1
RAD17/RAD50/RAP80	14	35.0	0	0.0	5	23.8	1	1.9	0	0.0	0	0.0	<0.001	20	11.1
10q23.31 (PTEN)	9	22.5	1	6.7	5	23.8	12	23.1	12	30.0	2	16.7	0.6	41	22.8
13q14.2 (RB1)	18	45.0	3	20.0	7	33.3	11	21.1	33	82.5	4	33.3	<0.001	76	42.2
17p13.1 (TP53)	20	50.0	6	40.0	12	57.1	18	34.6	20	50.0	4	33.3	0.4	80	44.4
17q21 (BRCA1)	17	42.5	2	12.5	11	52.4	11	21.2	7	17.5	3	25.0	0.01	51	28.3
13q12.1 (BRCA2)	12	30.0	3	18.8	5	23.8	8	15.4	21	52.5	3	25.0	0.01	52	28.9
4q31.21 (INPP4B)	16	40.0	1	11.1	1	4.8	6	11.5	4	10.0	1	8.3	0.002	29	16.1
<b>(b) Amplifications</b>															
17q12 (ERBB2)	9	22.5	2	13.3	12	57.1	9	17.3	15	37.5	2	16.7	0.007	49	27.2
12p12.1 (KRAS)	14	35.0	4	26.7	5	23.8	9	17.3	9	22.5	2	16.7	0.5	43	23.9
12q15 (MDM2)	6	15.0	1	6.7	4	19.0	12	23.1	18	45.0	1	8.3	0.01	42	23.3
8q24.21 (MYC)	26	65.0	7	46.7	10	47.6	18	34.6	29	72.5	3	25.0	0.002	93	51.7
10p11.23 (MAP3K8)	16	40.0	1	6.7	6	28.6	3	5.8	3	7.5	2	16.7	<0.001	31	17.2
10p11.22 (ZEB1)	16	40.0	1	6.7	5	23.8	3	5.8	3	7.5	2	16.7	<0.001	30	16.7
10p13 (FAM107B)	20	50.0	1	6.7	5	23.8	4	7.7	6	15.0	1	8.3	<0.001	37	20.6
<b>(c)</b>															
Average # of Gains	3943		1543		2970		2847		3885		2326				3192
Average # of Losses	4854		1906		3347		2560		4634		2891				3590
Total # of Aberrations	8797		3450		6317		5408		8519		5218				6782
Average # of Segments	194		143		223		150		222		139				183
Segment Length (kb)	14988		20421		13068		19424		13113		21062				15923
<b>(d)</b>															
% Tumor Cellularity (ASCAT)	52.5		38.4		38.2		53.0		50.2		38.1				48.4
% Tumor Purity (genoCNA)	75.0		78.0		79.5		68.0		68.0		70.0				71.0
<b>(e) Co-occurrence of 5q<sup>b</sup> loss with additional gene loss</b>															
	<i>N</i>	No.	%	<i>P</i> -Value <sup>c</sup>											
10q23.31 (PTEN) Loss	32	14	43.8	<0.001											
13q14.2 (RB1) Loss	32	26	81.3	<0.001											
17q21 (BRCA1) Loss	32	17	53.1	<0.001											
10p Amplicon	32	15	46.9	<0.001											
17p13.1 (TP53) Loss	32	19	59.4	0.08											
4q31.21 (INPP4B) Loss	32	16	50.0	<0.001											

Values are presented in 'Count (%)' format. Specific counts are given for individual deletions or co-deletions, with each sample only classified into one category. c Counts for average gains/losses for each subtype. Total number of aberrations is the sum of all individual gene gains and losses. Average segment number and length were calculated from the SWITCHdna generated segments for each sample within each subtype. d % Tumor Cellularity generated by ASCAT algorithm or genoCNA algorithm. e Rates of co-occurrence of 5q cluster loss with other gene alterations are shown (*N* refers to the number of total samples with 5q loss). Fisher's exact tests or Chi-square approximations were done to determine if the rates of occurrence, or co-occurrence, were at statistically significant levels

<sup>a</sup> Chi-square approximation

<sup>b</sup> RAD17+RAD50 loss OR RAD17+RAD50+RAP80 loss

<sup>c</sup> Fisher's exact test

**Table 2** Frequency of copy number alterations data for the Jonsson dataset [14] for selected (a) deletions, (b) amplifications, (c) average number of changes, and (d) co-occurrences

	Jonsson												All ( <i>n</i> = 356)		
	Basal ( <i>n</i> = 61)		Claudin ( <i>n</i> = 43)		Her2 ( <i>n</i> = 46)		LumA ( <i>n</i> = 117)		LumB ( <i>n</i> = 55)		Normal-like ( <i>n</i> = 34)		<i>P</i> -Value <sup>a</sup>	No.	%
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%			
<b>(a) Deletions</b>															
No 5q Genes Lost	15	24.6	29	67.4	28	58.3	107	90.7	44	80.0	26	76.5	<0.001	249	69.9
5q13.2 (RAD17)	9	14.8	4	9.3	3	6.3	3	2.5	5	9.1	3	8.8	<0.001	27	7.6
5q31.1 (RAD50)	4	6.6	0	0.0	2	4.2	3	2.5	3	5.5	1	2.9	<0.001	13	3.7
5q35.2 (RAP80)	0	0.0	0	0.0	0	0.0	2	1.7	0	0.0	1	2.9	<0.001	3	0.8
RAD17/RAD50	16	26.2	2	4.7	12	25.0	2	1.7	2	3.6	2	5.9	<0.001	36	10.1
RAD17/RAP80	3	4.9	1	2.3	0	0.0	0	0.0	0	0.0	0	0.0	<0.001	4	1.1
RAD50/RAP80	2	3.3	0	0.0	1	2.1	0	0.0	1	1.8	0	0.0	<0.001	4	1.1
RAD17/RAD50/RAP80	12	19.7	7	16.3	2	4.2	1	0.8	0	0.0	1	2.9	<0.001	23	6.5
10q23.31 (PTEN)	21	34.4	10	23.3	9	18.8	14	11.9	17	30.9	2	5.9	<0.001	73	20.5
13q14.2 (RB1)	33	54.1	16	37.2	13	27.1	32	27.1	31	56.4	8	23.5	<0.001	133	37.4
17p13.1 (TP53)	21	34.4	6	14.0	17	35.4	33	28.0	19	34.5	11	32.4	0.2	107	30.1
17q21 (BRCA1)	19	31.1	6	14.0	7	15.2	12	10.3	5	9.1	5	14.7	<0.001	54	15.2
13q12.1 (BRCA2)	22	36.1	13	30.2	12	26.1	26	22.2	30	54.5	7	20.6	<0.001	110	30.9
4q31.21 (INPP4B)	29	47.5	14	32.6	16	34.8	13	11.1	11	20.0	2	5.9	<0.001	85	23.9
<b>(b) Amplifications</b>															
17q12 (ERBB2)	9	14.8	5	11.6	31	64.6	23	19.5	17	30.9	6	17.6	<0.001	91	25.6
12p12.1 (KRAS)	12	19.7	7	16.3	3	6.5	3	2.5	10	18.2	0	0.0	<0.001	35	9.8
12q15 (MDM2)	2	3.3	4	9.3	7	15.2	13	11.1	17	30.9	0	0.0	<0.001	43	12.1
8q24.21 (MYC)	42	68.9	19	44.2	22	47.8	48	41.0	44	80.0	14	41.2	<0.001	189	53.1
10p11.23 (MAP3K8)	16	26.2	8	18.6	7	14.6	6	5.1	8	14.5	1	2.9	0.001	46	12.9
10p11.22 (ZEB1)	16	26.2	8	18.6	8	16.7	6	5.1	8	14.5	1	2.9	<0.001	47	13.2
10p13 (FAM107B)	29	47.5	10	23.3	6	12.5	6	5.1	11	20.0	3	8.8	<0.001	65	18.3
<b>(c)</b>															
Average # of Gains	2853		2169		2481		1963		3281		1469		4675		
Average # of Losses	5089		3171		3523		2430		3872		2395		6597		
Total # of Aberrations	7942		5341		6003		4393		7153		3864		11272		
Average # of Segments	167		130		134		97		129		93		122		
Segment Length (kb)	16522		21289		20573		28542		21462		29841		22610		
<b>(d) Co-occurrence of 5q<sup>b</sup> loss with additional gene loss</b>															
	<i>N</i>	No.		%		<i>P</i> -Value <sup>c</sup>									
10q23.31 (PTEN) Loss	67	28		41.8		<0.001									
13q14.2 (RB1) Loss	67	39		58.2		<0.001									
17q21 (BRCA1) Loss	67	24		35.8		<0.001									
10p Amplicon	67	15		22.4		0.02									
17p13.1 (TP53) Loss	67	29		43.3		0.01									
4q31.21 (INPP4B) Loss	67	31		46.3		<0.001									

Values are presented in Count (%) format. Specific counts are given for individual deletions or co-deletions, with each sample only classified into one category. c Counts for average gains/losses for each subtype. Total number of aberrations is the sum of all individual gene gains and losses. Average segment number and length were calculated from the SWITCHdna generated segments for each sample within each subtype. d Rates of co-occurrence of 5q cluster loss with other gene alterations are shown (*N* refers to the number of total samples with 5q loss). Fisher's exact tests or Chi-square approximations were done to determine if the rates of occurrence, or co-occurrence, were at statistically significant levels

<sup>a</sup> Chi-square approximation

<sup>b</sup> RAD17+RAD50 loss OR RAD17+RAD50+RAP80 Loss

<sup>c</sup> Fisher's exact test

**Table 3** Comparison of Jonsson et al. copy number based classifications versus intrinsic subtypes

		Jonsson subtypes					
		Amplifier	Luminal-complex	Mixed	17q12	Luminal-simple	Basal-complex
PAM50+Claudin low subtypes	Basal	7	4	4	3	1	42
	LumA	18	42	15	10	31	2
	Claudin	11	4	8	4	3	13
	LumB	11	36	3	2	1	2
	Her2	1	9	3	30	0	5
	Normal-like	4	10	5	2	10	3

Subtype classifications using both the original labels in the Jonsson dataset [14] using copy number defined subtypes, versus PAM50 plus Claudin-low gene expression subtypes is shown. *P* value determined by Chi-square approximation

*P*-value < 0.001

subtypes (*P* value < 0.001, Table 3); importantly, there was high overlap between our Basal-like subtype and their Basal-complex phenotype, both of which showed the frequent loss of 5q11–35 and amplification of 10p.

#### Increased genomic instability of tumors associated with loss of specific regions/genes

To objectively assess “genomic instability”, we calculated a loss/normal/gain value for every gene using the SWITCHdna assigned copy number states, and calculated the levels of genomic instability by subtype using the average number of gains/losses per sample on a gene by gene basis. The Basal-like subtype was the most prone to aberrations, while the Claudin-low and Luminal A subtypes showed the lowest

number of gene-based CNA (Table 1c). To control for a large number of genes being gained or lost by a large single genomic aberration event (i.e., whole chromosome loss), we also calculated the average number of SWITCHdna-defined segments and their length for each subtype, as more genomic breaks will result in more segments. The subtypes that had greater numbers of gene aberrations were also the same ones that had more SWITCHdna segments of shorter average length (Table 1c). Thus, the increased number of aberrant gene-based events in the copy number unstable subtypes was due to more frequent aberrations in the genome, rather than as a large number of genes gained or lost by a few large-in-size aberration events.

Tumors with loss of *PTEN*/10q23.31, *RBI*/13q14.2, or *TP53*/17p13.1, or amplification of the 10p region were also

**Table 4** Examination of possible correlations between the specific CNA and overall genomic instability

Average amount of CNA by event and class in combined UNC/Norway dataset	Average gain	Average loss	Average total
All samples mean ( <i>n</i> = 180)	3,192	3,590	6,783
All samples median ( <i>n</i> = 180)	2,790	3,230	6,560
RAD17 + RAD50 ( <i>n</i> = 12)	3,948 ( <i>P</i> = 0.1)	5,614 ( <i>P</i> = 0.0004*)	9,562 ( <i>P</i> = 0.003*)
RAD17 + RAD50 +/- RAP80 ( <i>n</i> = 32)	4,451 ( <i>P</i> = 0.0002*)	6,681 ( <i>P</i> < 0.00001*)	11,131 ( <i>P</i> < 0.00001*)
RAD17 + RAD50 + RAP80 ( <i>n</i> = 20)	4,752 ( <i>P</i> = 0.0002*)	7,320 ( <i>P</i> < 0.00001*)	12,073 ( <i>P</i> < 0.00001*)
Other ( <i>n</i> = 148)	2,920	2,922	5,842
10q23.31 ( <i>PTEN</i> ) loss ( <i>n</i> = 31)	4,202 ( <i>P</i> = 0.009*)	6,169 ( <i>P</i> < 0.00001*)	10,371 ( <i>P</i> < 0.00001*)
No 10q23.31 ( <i>PTEN</i> ) loss ( <i>n</i> = 149)	2,982	3,054	6,036
13q14.2 ( <i>RB1</i> ) loss ( <i>n</i> = 66)	4,127 ( <i>P</i> = 0.00002*)	5,638 ( <i>P</i> < 0.00001*)	9,765 ( <i>P</i> < 0.00001*)
No 13q14.2 ( <i>RB1</i> ) loss ( <i>n</i> = 114)	2,652	2,404	5,056
17p13.1 ( <i>TP53</i> ) loss ( <i>n</i> = 80)	3,900 ( <i>P</i> = 0.00008*)	4,8578 ( <i>P</i> < .00001*)	8,757 ( <i>P</i> < 0.00001*)
No 17p13.1 ( <i>TP53</i> ) loss ( <i>n</i> = 100)	2,627	2,576	5,203
10p Amplicon ( <i>n</i> = 34)	5,016 ( <i>P</i> < 0.00001*)	5,232 ( <i>P</i> = 0.0002*)	10,248 ( <i>P</i> < 0.00001*)
No 10p Amplicon ( <i>n</i> = 146)	2,768	3,208	5,975

The average numbers of CNAs for gains, losses, or both, are shown for the entire dataset and within sets of tumors with a given copy number alteration (5q, *PTEN*/10q23.31, *RBI*/13q14.2, *TP53*/17p13.1, and 10p). A Wilcoxon-rank sum test was performed to see if the rate of copy number aberration between each group (Pairwise: Aberration vs. Other, or No Aberration) was significantly different (\*)

found to have high rates of total gene-based CNA compared to tumors without loss of these genes (Table 4). Loss of 5q11–35 was also associated with the highest numbers of CNA, with the greatest instability seen when all three DNA repair genes were lost.

Low expression of genes residing in Basal-like regions correlates with poor survival and predicts therapeutic response

To determine if these DNA loss events also impacted gene function, we determined whether the mRNA levels of candidate genes contained within these regions correlated with DNA loss. The expression of ten genes selected based on their associations with the basal-like subtype, or breast cancer in general, was evaluated. Most showed significantly lower mRNA expression when the genomic DNA was lost including *RAD17*, *RAD50*, *RAP80*, *MSH3*, *RB1*, *PTEN*, *BRCA1*, and *INPP4B* (Fig. 2); these data suggest that these losses have functional consequences (noting that only *TP53* and *BRCA2* did not show in *cis* correlation between expression and copy number). It is also of note that *MSH3* (a gene involved in DNA mismatch repair), located within the 5q11–35 loss region (between *RAD17* and *RAD50*, Fig. 1h), and it also showed reduced mRNA expression when lost and low expression within Basal-like tumors in general (Figs. 2, 3e). In addition, the mRNA expression levels of *RAD17*, *RAD50*, *MSH3*, *RAP80*, *INPP4B*, and *PTEN* were lowest in the Basal-like subtype (Fig. 3, UNC337 expression dataset [5]); thus loss of 5q11–35 likely affects multiple aspects of DNA repair.

Using patient survival data from two additional data sets containing gene expression data (UNC337 [5] and NKI295 [28]), Kaplan–Meier analysis showed that the low average expression of *RAD17*+*RAD50* was associated with worse outcomes compared to high expression (Fig. 4a). A similar trend was observed with *INPP4B*, mirroring previous observations (Fig. 4b) [24]. *RAD17*+*RAD50* expression was also examined for treatment effects using the Hess et al. [29] data set, which examined T/FAC neoadjuvant chemotherapy responsiveness across 130 breast cancer patients. Low expression of *RAD17*+*RAD50* was correlated with pathological complete response (pCR) (ANOVA *P* value <0.0001). This finding may be due to the association between low expression of *RAD17*+*RAD50* and Basal-like tumors, as Basal-like tumors have also been shown to have high neoadjuvant chemotherapy pCR rates [30, 31].

Knockdown of *RAD17*±*RAD50* affects sensitivity to chemotherapeutics and BRCA1 foci formation

Given the involvement of *RAD17*, *RAD50*, and *RAP80* in the BRCA1-DNA repair pathway, we determined whether disruption of these genes via RNAi knockdown would lead

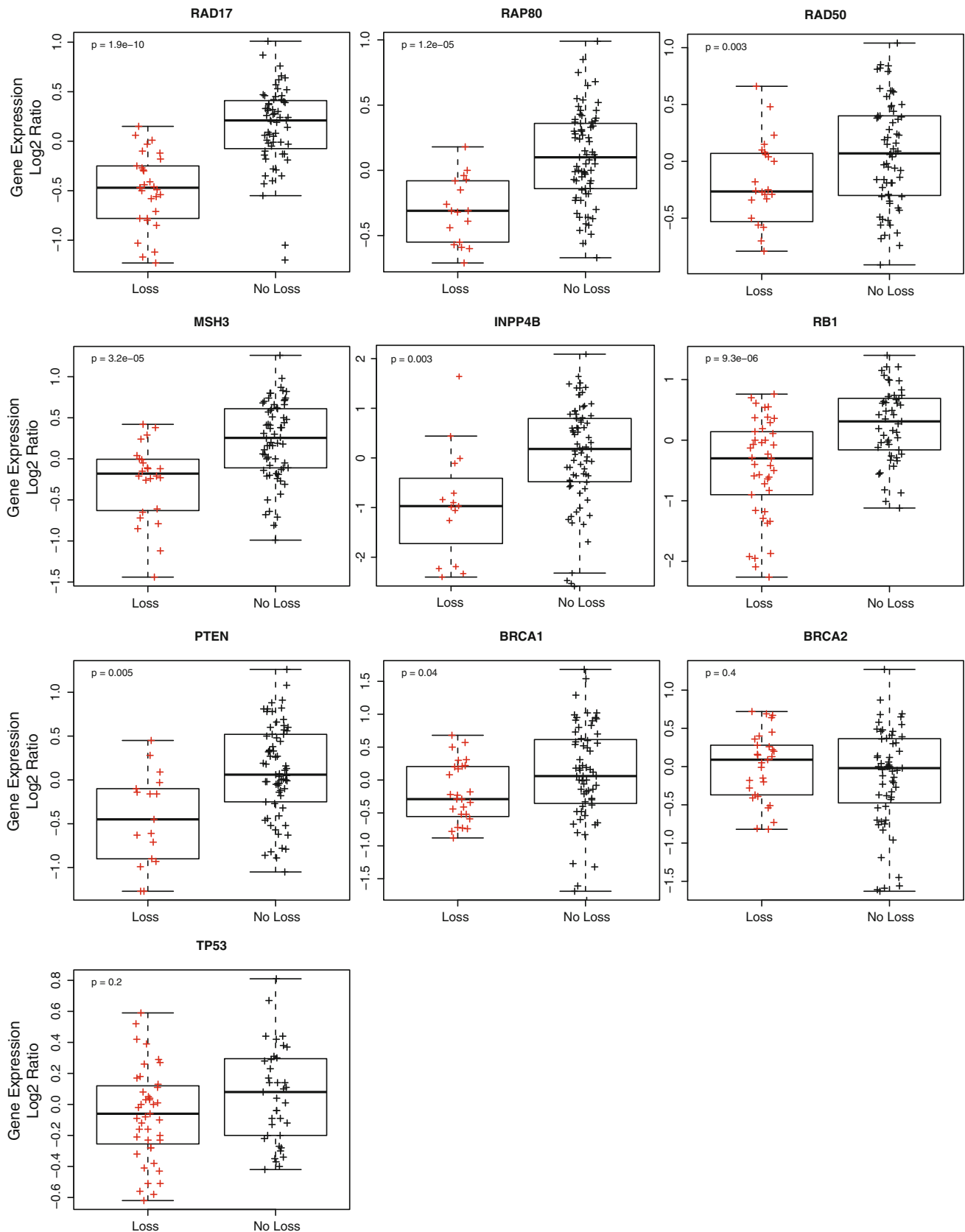
to changes in sensitivity to drugs whose mechanism of action has already been linked to BRCA1 loss like carboplatin/cisplatin [32, 33] and PARP inhibitors [34, 35]. *RAD17* was stably knocked down with shRNA in the HME-CC cell line (an hTERT-immortalized Human Mammary Epithelial Cell) [36] and knockdown was confirmed by Western blotting (Fig. 5a). HME-CC cells with *RAD17* knockdown exhibited increased sensitivity to ABT-888 (PARPi) and carboplatin (Fig. 5c). No difference in paclitaxel sensitivity was observed, which was used as a non-DNA-damaging agent control. A *RAD50* knockdown line did not exhibit any change in sensitivity to ABT-888 and had a paradoxical increase in resistance to carboplatin. We next emulated the most common in vivo co-occurring loss by generating a double knockdown of *RAD17* and *RAD50*, which showed the greatest increased sensitivity to ABT-888 and carboplatin (Fig. 5c). Similar results were observed when this experiment was repeated in ME16C cells, a second hTERT-immortalized human mammary epithelial cell line (Supplemental Fig. 2).

In order to assess the effects of *RAD17*/*RAD50* loss on BRCA1-dependent DNA repair, we performed a DNA repair foci formation assay on the control and *RAD17*+*RAD50* double knockdown line. Using anti-BRCA1 protein immunofluorescence, and automated foci counting within geminin-positive cells, we observed a significant decrease in the number of BRCA1-containing DNA repair foci in the double knockdown line when treated with ionizing radiation or ABT888 versus control (Fig. 6); cells were simultaneously stained for geminin in order to control for differences in proliferation as described by Graeser et al. [37, 38]. These data suggest that loss of *RAD17* and/or *RAD50* may impair BRCA1 function, and could contribute to increased sensitivity to DNA-damaging agents.

## Discussion

The presence of distinct breast cancer expression subtypes suggests different underlying genetic events may be driving each subtype. To address this hypothesis, we used 180 diverse tumors and performed supervised analyses of their tumor DNA copy number landscape and identified subtype-specific copy number events. Many studies have identified numerous regions of gain and loss in human breast tumors [9, 10, 14, 23, 39]; however, most did not specifically search for regions uniquely associated with specific intrinsic subtypes. Some previous attempts were made to identify basal-like specific CNA [10, 22] and we observed a number of the same findings. We take these previous findings as validation of our identified regions, and we build and expand upon these here, along with the addition of functional studies.





**Fig. 2** Gene expression values for *RAD17*, *RAD50*, *RAP80*, *MSH3*, *BRCA1*, *BRCA2*, *PTEN*, *RB1*, *TP53*, and *INPP4B* in the UNC-Norway dataset (n = 180) separated by copy number status (DNA copy number loss vs no loss). P values determined by ANOVA test

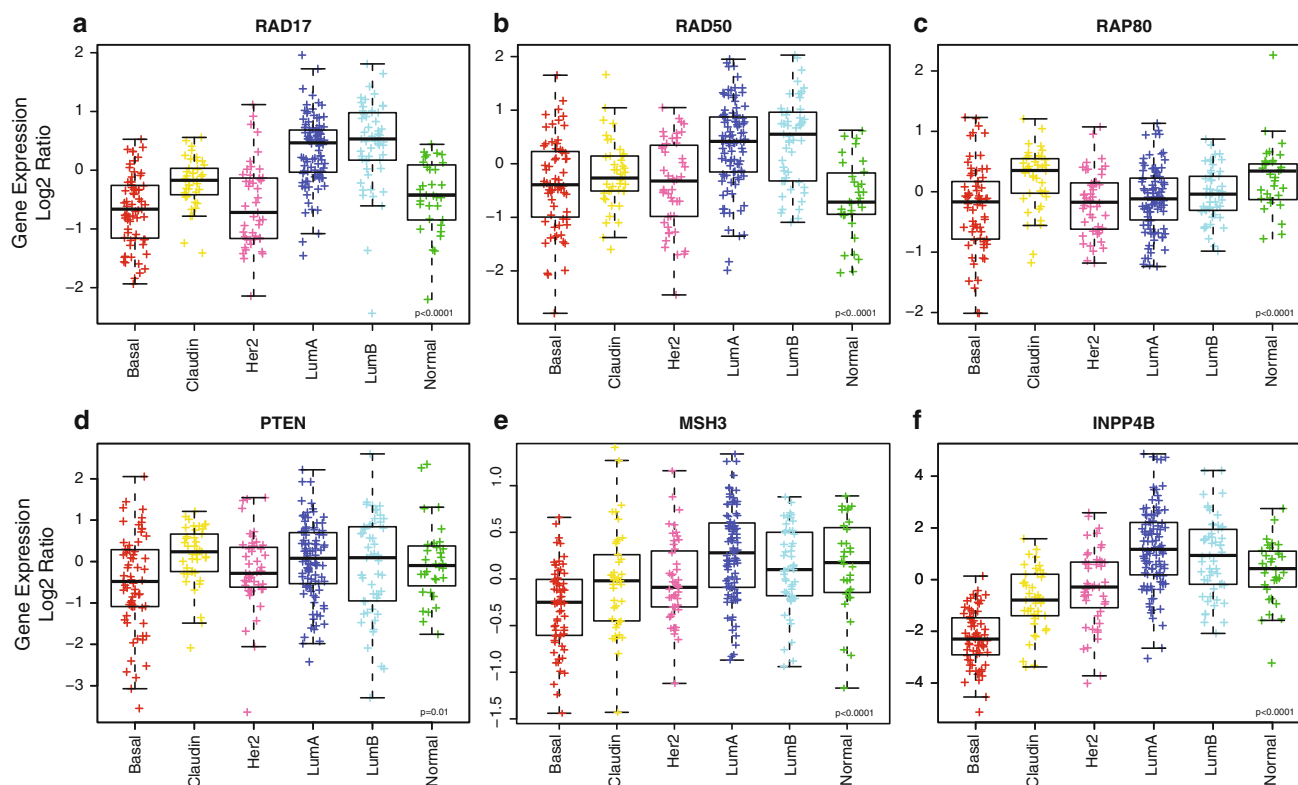
Overall, we identified many subtype-specific CNA and validated these findings on a second, independent dataset. Here we have focused on the Basal-like subtype, which showed by far the greatest number of subtype-specific CNA and were the most genomically unstable as determined by the sheer number of CNA, a feature which has been observed in the past [9]. Basal-like tumors also showed consistent loss of 4q (which harbors *INPP4B* and *FBXW7*), and 5q11–35, which contains many DNA repair genes. Basal-like tumors are known to be associated with BRCA1-pathway dysfunction in that 80–90% of BRCA1 mutation carriers, if and when they develop breast cancer, develop Basal-like tumors [3, 40, 41]; however, in most sporadic Basal-like tumors, the *BRCA1* gene appears normal in sequence [42]. The loss of 5q11–35 may provide an alternative means to impair BRCA1-pathway function and explain why despite many Basal-like patients having normal BRCA1 gene/protein, high levels of genomic instability and a “BRCAness” phenotype are observed in Basal-like tumors. Previous evidence indicates a link between genes involved in BRCA1 DNA damage control and genes that are deleted and downregulated in Basal-like cancers, lending further credence to our hypothesis [26].

In order to expand our understanding of the relationship between the Basal-like subtype and impaired BRCA1-pathway function, we pursued functional studies by RNAi-

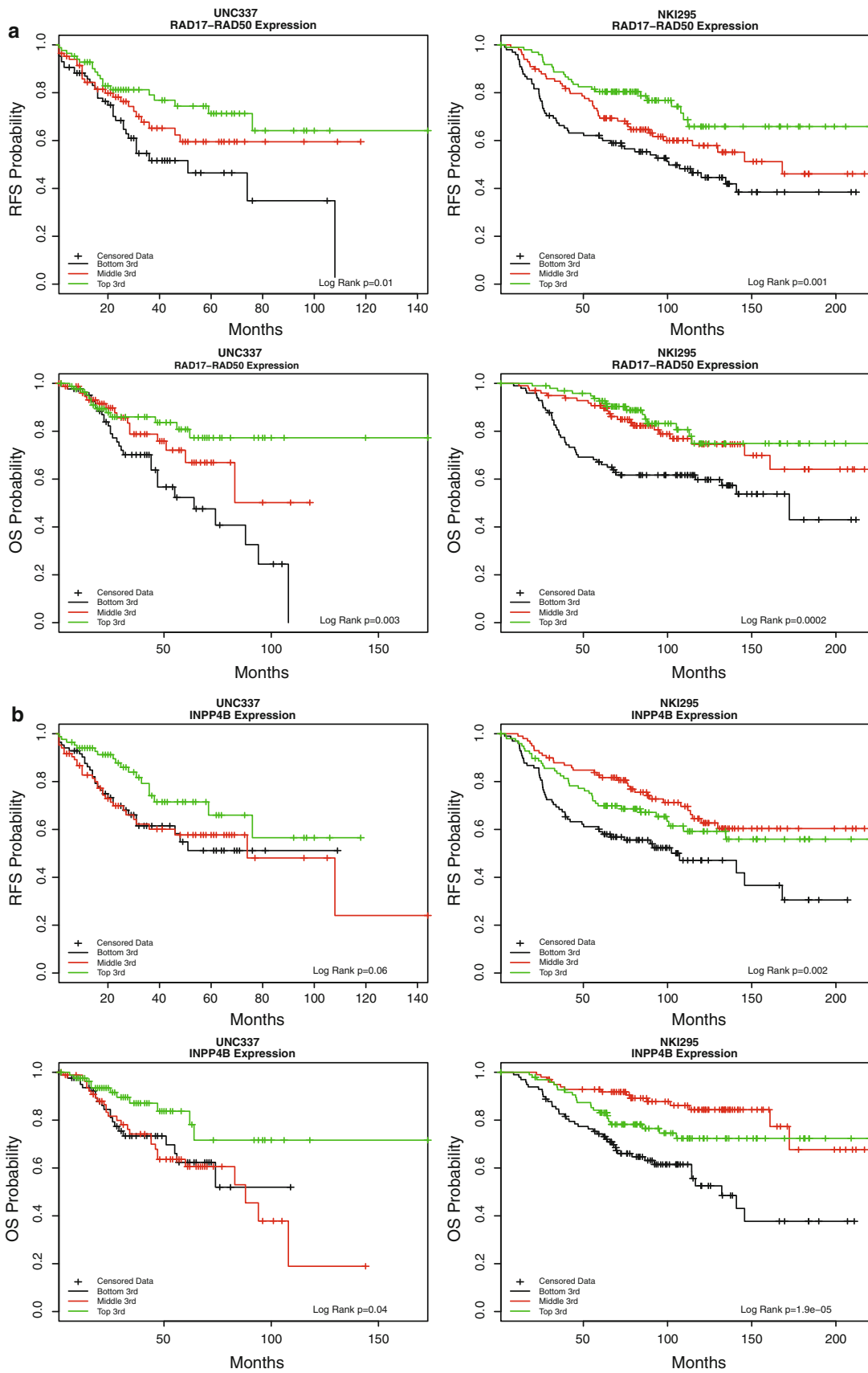
**Fig. 4** Survival analysis according to expression of RAD17+RAD50 and INPP4B. Patients in the UNC337 and NKI295 data sets were ranked ordered by average gene expression values of **a** *RAD17+RAD50* combined, or **b** *INPP4B*. The patients were split into thirds based upon rank order expression values and Kaplan–Meier analysis was done on the three groups to examine trends in relapse-free survival and overall survival. *P* values determined by log-rank test

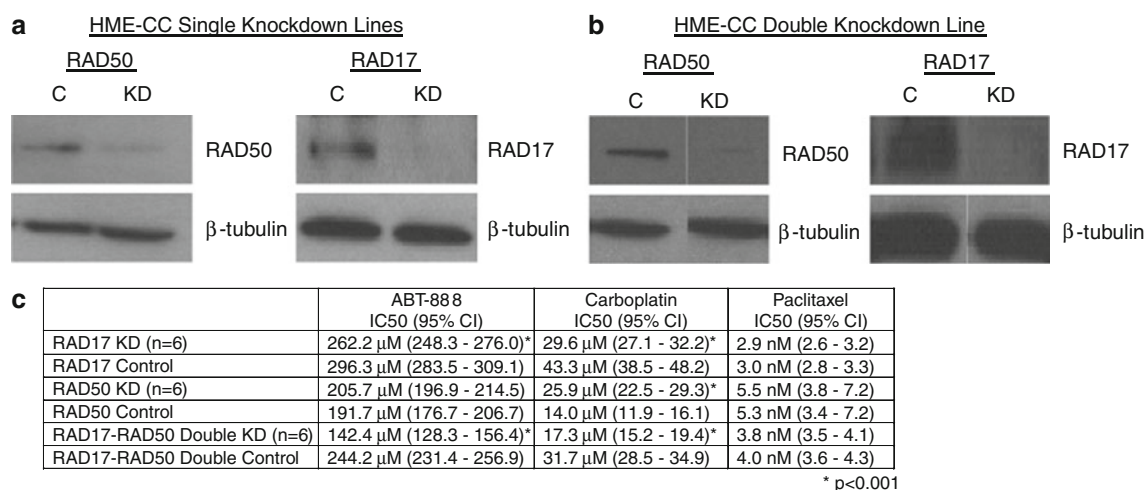
mediated knockdown of two members of the pathway, RAD17 and RAD50, in order to emulate the genomic losses observed in tumors. Besides being members of the BRCA1-pathway, others have highlighted these genes for their possible Basal-like association, but without functional studies [10, 27]. We show here that genetic ablation of these genes results in impaired DNA repair and increased drug sensitivity, and furthermore, deletion of RAD17 and RAD50 in yeast has also been shown to result in increased sensitivity to DNA-damaging agents including platinum drugs (<http://fitdb.stanford.edu>) [43]; these data highlight that there is an evolutionarily conserved role for these genes in DNA repair.

By building upon the discovery of the subtype association and the deletion phenotypes in yeast, we propose a role in DNA repair function for the 5q11–35 region. The drug sensitivity assays show the importance of these genes in DNA damage sensitivity and the foci formation



**Fig. 3** ANOVA boxplots for individual genes that are commonly lost in Basal-like cancers according to intrinsic subtype determined using the UNC337 sample set. *P* values were determined by 2-way ANOVA. **a** *RAD17*, **b** *RAD50*, **c** *RAP80*, **d** *PTEN*, **e** *MSH3*, and **f** *INPP4B*





**Fig. 5** RNAi knockdown experiments in an immortalized HMEC (BABE cell line). Western blot analysis showing reduction of RAD17 and RAD50 protein expression in HME-CC **a** single, or **b** double RNAi knockdown lines. (KD knockdown line, C vector control line). Tubulin staining was performed as a loading control. **c** Estimated

IC50 with 95% CI for ABT-888, Carboplatin, and Paclitaxel based on mitochondrial dye-conversion assay. Results are based on the average of two experiments per condition, each done in triplicate, with knockdown-control pairs with significant differences in IC50 are designated with a \*

experiments show that their function is mediated through BRCA1. In addition, from the combination of our genomic analyses and functional data, it is our hypothesis that the somatic loss of *RAD17*, *RAD50*, and/or *RAP80* leads to impaired BRCA1-pathway function, impaired homologous recombination mediated DNA repair, and thus, contributes to overall genomic instability.

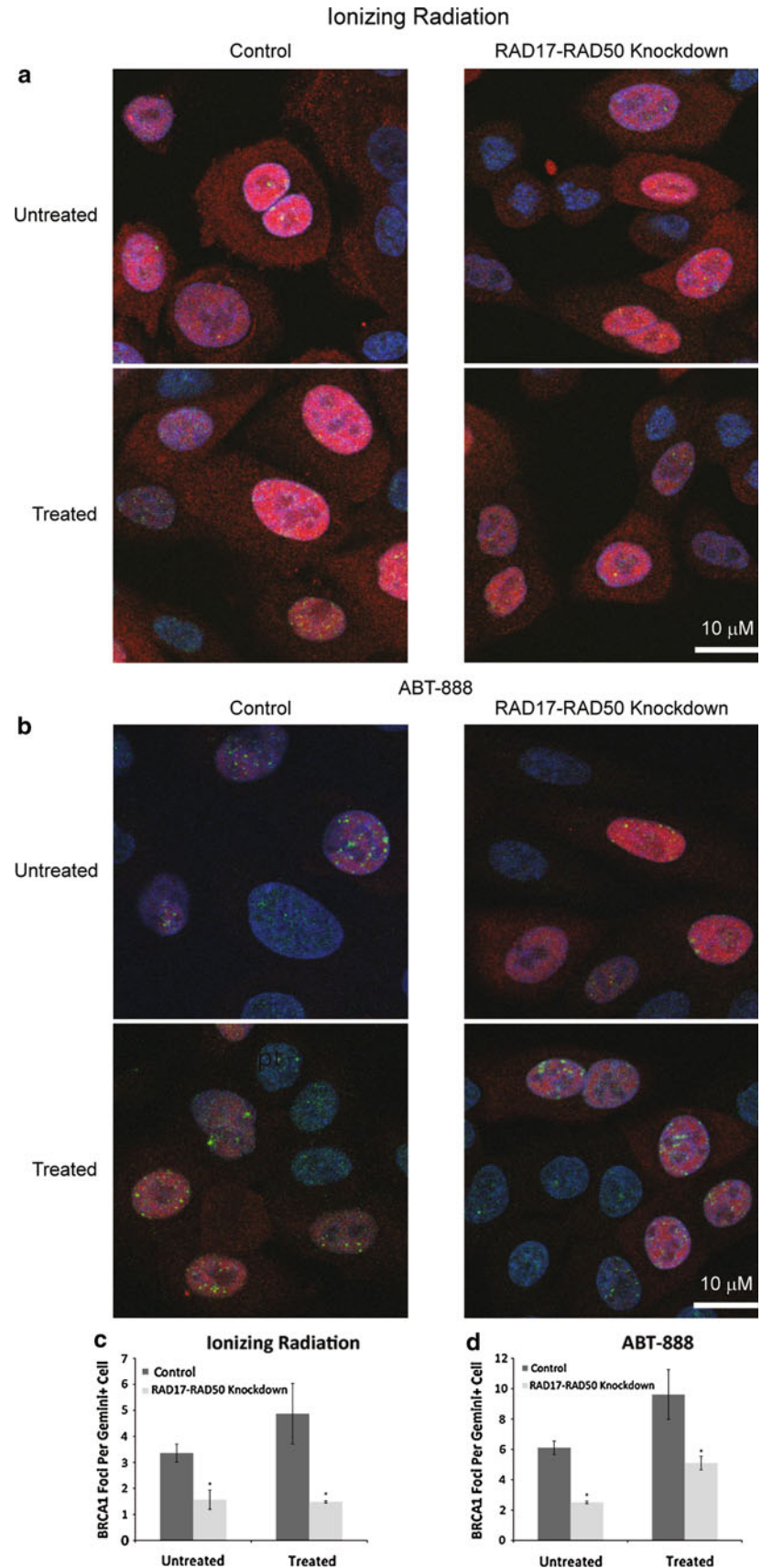
There are, however, two caveats to these analyses and our hypothesis. First, the 5q11–35 loss is a large region that typically involves >100 genes, therefore, we cannot definitively say that loss of these three genes is the target of this deletion, or that these three genes are the most important targeted genes of this region. Second, a high frequency of co-occurrence with other DNA chromosomal losses happens in tumors with 5q11–35 loss; for example, in ~80% of tumors with 5q11–35 loss, *RBI/13q14.2* DNA loss also occurs (and by itself is associated with increased genomic instability). In addition, ~60% of these tumors show *TP53/17p13.1* loss (Table 1, 2). The co-occurrence of 5q11–35 loss with *RBI* and *TP53* loss are likely causative events in Basal-like carcinogenesis (the latter two being corroborated by mouse studies) [44–46]. Given the high co-occurrence of chromosome region losses that are not physically linked, it is impossible to say which one is the cause of the genomic instability. However, our hypothesis is that each of these regions harbors genes needed for maintenance of the genome and that the combinatorial loss of 2–3 of these regions is what results in the genomic instability phenotype seen in Basal-like breast cancers. In this article, we examine DNA losses, but do note that it is possible that loss of these same genes could also occur via methylation, altered microRNA regulation,

and/or somatic mutation, although the last of these has yet to be found when searching current somatic mutation databases for *RAD17/RAD50/RAP80*. Preliminary sequence analysis of *RAD17* and *RAD50* (data not shown), as well as evaluation of previous breast cancer sequencing efforts [47] and the COSMIC database [48], revealed few, if any, somatic variants/mutations in these two genes, which is consistent with the finding that loss of any one gene is rarely seen; thus, if loss of two or more genes is the target of this CNA, then somatic mutation of any one gene would not impart a selective tumorigenic advantage. Therefore, these data suggest that the target of 5q11–35 loss is two or more genes in this region, with loss of *RAD17* and *RAD50* likely contributing to genomic instability.

## Conclusions

The gene expression-defined intrinsic subtypes of breast cancer are mirrored by DNA copy number changes. The Basal-like subtype is the most distinct in the copy number landscape world, and these subtype-associated CNA have clinical implications. If 5q11–35 loss results in impaired homologous recombination mediated DNA repair, as was suggested by our in vitro studies and in vivo correlates, then the loss of this region may sensitize tumors to specific classes of DNA-damaging agents. Based upon BRCA1 studies in vitro [49, 50] and in vivo [32, 34], these drugs could include PARP inhibitors and cis/carboplatin. Loss of *RAD17+RAD50* (mRNA and/or genomic DNA) may thus be a biomarker of chemotherapy responsiveness, which is supported by our finding of an association for predicting a

**Fig. 6** BRCA1-mediated DNA repair foci formation assay. **a** Representative images of BRCA1 foci formation in RAD17-RAD50 double knockdown cells and control cells after treatment with 2.5 Gy of ionizing irradiation and 20 min recovery (ionizing radiation), or no treatment (untreated). **b** Representative images of BRCA1 foci formation in RAD17-RAD50 double knockdown cells and control cells with 200  $\mu$ M ABT-888 (ABT-888), or no treatment (untreated). *Green channel* BRCA1, *Red channel* Geminin, *Blue channel* DAPI images. All images were taken with a 63 $\times$  objective and post processed to 300% of their original size. Automated BRCA1 foci counting results from each cell line for **c** ionizing radiation and **d** ABT-888 treatment. *Error bars* represent 95% confidence intervals (\* $P < 0.05$  of knockdown relative to control). *P* values were calculated from *t* tests comparing foci counts in treated double knockdown cells versus treated control cells or untreated double knockdown cells versus untreated control cells



likelihood of achieving a pathological complete response. We hypothesize that the loss of these DNA repair genes and the 5q11–35 region, contributes to genomic instability and mutability, ultimately causing high proliferation rates and aggressive behaviors. Our integrated studies of gene expression and genomic DNA copy number have identified important pathway-based determinants of Basal-like cancers and a possible therapeutic biomarker.

All relevant gene expression and copy number data new to this manuscript can be found in the GEO database under series GSE10893.

**Acknowledgments** This study was supported by funds from the NCI Breast SPORE program (P50-CA58223), RO1-CA138255, T32-GM008719, F30-ES018038, R03-CA132143, P50-CA125183, the Breast Cancer Research Foundation, the EIF-Lee Jeans Translational Research Fund, and the V Foundation for Cancer Research. We thank the UNC CICBDD for ABT-888, which is directed by Stephen Frye, and compound provision is managed by Jian Jin.

**Conflict of interest** CMP and JSP are listed as inventors on the PAM50 intrinsic subtyping pending patent application. CMP owns stock in Bioclassifier, LLC, which has licensed the rights to the PAM50 intrinsic subtyping assay. A research only and publicly available version of the PAM50 assay was used in this manuscript. The remaining authors declare that they have no competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA (2000) Molecular portraits of human breast tumours. *Nature* 406(6797): 747–752
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale A-L (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98(19):10869–10874. doi:10.1073/pnas.191367098
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 100: 8418–8423
- Hu Z, Fan C, Oh D, Marron JS, He X, Qaqish B, Livasy C, Carey L, Reynolds E, Dressler L, Nobel A, Parker J, Ewend M, Sawyer L, Wu J, Liu Y, Nanda R, Tretiakova M, Orrico A, Dreher D, Palazzo J, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush J, Ellis M, Olopade O, Bernard P, Perou C (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7(1):96
- Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JJ, He X, Perou CM (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12(5):R68
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS (2009) Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J Clin Oncol*. doi: 10.1200/JCO.2008.18.1370
- Pinkel D, Albertson DG (2005) Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 37(Suppl):S11–S17
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom 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, Pet-hiyagoda 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(5853):1108–1113
- Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10(6):529–541
- Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Bousard T, Lonning PE, Tibshirani R, Borresen-Dale AL, Pollack JR (2006) Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* 45(11):1033–1040
- Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Borresen-Dale AL, Kristensen VN (2010) Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA*. doi: 10.1073/pnas.1009843107
- Russnes HG, Vollan HK, Lingjaerde OC, Krasnitz A, Lundin P, Naume B, Sorlie T, Borgen E, Rye IH, Langerod A, Chin SF, Teschendorff AE, Stephens PJ, Maner S, Schlichting E, Baum-busch LO, Karesen R, Stratton MP, Wigler M, Caldas C, Zetterberg A, Hicks J, Borresen-Dale AL (2010) Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med* 2(38):38ra47
- 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, Oberkell 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(7291):999–1005
- Jonsson G, Staaf J, Vallon-Christersson J, Ringner M, Holm K, Hegardt C, Gunnarsson H, Fagerholm R, Strand C, Agnarsson BA, Kilpivaara O, Luts L, Heikkila P, Aittomaki K, Blomqvist C, Loman N, Malmstrom P, Olsson H, Johannsson OT, Arason A, Nevanlinna H, Barkardottir RB, Borg A (2010) Genomic subtypes of breast cancer identified by array-comparative genomic

- hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res* 12(3):R42
15. Andrews DWK (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61:821–856
  16. Bai J, Perron B (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66:47–78
  17. R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
  18. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B (Methodological)* 57(1):289–300
  19. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa JL, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavare S, Brenton JD, Ylstra B, Caldas C (2007) High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol* 8(10):r215. doi:[10.1186/gb-2007-8-10-r215](https://doi.org/10.1186/gb-2007-8-10-r215)
  20. Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, Minna JD, Pollack JR (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* 4(7):e6146. doi:[10.1371/journal.pone.0006146](https://doi.org/10.1371/journal.pone.0006146)
  21. Bergamaschi A, Kim YH, Kwei KA, La Choi Y, Bocanegra M, Langerod A, Han W, Noh DY, Huntsman DG, Jeffrey SS, Borresen-Dale AL, Pollack JR (2008) CAMK1D amplification implicated in epithelial-mesenchymal transition in basal-like breast cancer. *Mol Oncol* 2(4):327–339
  22. Adelaide J, Finetti P, Bekhouche I, Repellini L, Geneix J, Sir-coulomb F, Charafe-Jauffret E, Cervera N, Desplans J, Parzy D, Schoenmakers E, Viens P, Jacquemier J, Birnbaum D, Bertucci F, Chaffanet M (2007) Integrated profiling of basal and luminal breast cancers. *Cancer Res* 67(24):11565–11575. doi:[10.1158/0008-5472.can-07-2536](https://doi.org/10.1158/0008-5472.can-07-2536)
  23. Haverty PM, Fridlyand J, Li L, Getz G, Beroukheim R, Lohr S, Wu TD, Cavet G, Zhang Z, Chant J (2008) High-resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes Chromosomes Cancer* 47(6):530–542
  24. Gewinner C, Wang ZC, Richardson A, Teruya-Feldstein J, Etemadmoghadam D, Bowtell D, Barretina J, Lin WM, Rameh L, Salmena L, Pandolfi PP, Cantley LC (2009) Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* 16(2):115
  25. Fedele CG, Ooms LM, Ho M, Vieuxseux J, O'Toole SA, Millar EK, Lopez-Knowles E, Sriratana A, Gurung R, Baglietto L, Giles GG, Bailey CG, Rasko JE, Shields BJ, Price JT, Majerus PW, Sutherland RL, Tiganis T, McLean CA, Mitchell CA Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc Natl Acad Sci USA* 107(51):22231–22236
  26. Natrajan R, Weigelt B, Mackay A, Geyer FC, Grigoriadis A, Tan DS, Jones C, Lord CJ, Vatcheva R, Rodriguez-Pinilla SM, Palacios J, Ashworth A, Reis-Filho JS (2010) 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* 121(3):575–589
  27. Johannsdottir HK, Jonsson G, Johannsdottir G, Agnarsson BA, Eerola H, Arason A, Heikkila P, Egilsson V, Olsson H, Johannsson OT, Nevanlinna H, Borg A, Barkardottir RB (2006) Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. *Int J Cancer* 119(5):1052–1060
  28. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347(25):1999–2009
  29. Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, Booser D, Theriault RL, Buzzard AU, Dempsey PJ, Rouzier R, Sneige N, Ross JS, Vidaurre T, Gomez HL, Hortobagyi GN, Puzstai L (2006) Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 24(26):4236–4244
  30. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM (2007) The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 13(8):2329–2334. doi:[10.1158/1078-0432.ccr-06-1109](https://doi.org/10.1158/1078-0432.ccr-06-1109)
  31. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, Hess KR, Stec J, Ayers M, Wagner P, Morandi P, Fan C, Rabiul I, Ross JS, Hortobagyi GN, Puzstai L (2005) Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 11(16):5678–5685
  32. Silver DP, Richardson AL, Eklund AC, Wang ZC, Szallasi Z, Li Q, Juul N, Leong CO, Calogrias D, Buraimoh A, Fatima A, Gelman RS, Ryan PD, Tung NM, De Nicolo A, Ganesan S, Miron A, Colin C, Sgroi DC, Ellisen LW, Winer EP, Garber JE (2010) Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. *J Clin Oncol* 28(7):1145–1153
  33. Chang HR, Glaspy J, Allison MA, Kass FC, Elashoff R, Chung DU, Gornbein J (2010) Differential response of triple-negative breast cancer to a docetaxel and carboplatin-based neoadjuvant treatment. *Cancer* 116(18):4227–4237
  34. Donawho CK, Luo Y, Luo Y, Penning TD, Bauch JL, Bouska JJ, Bontcheva-Diaz VD, Cox BF, DeWeese TL, Dillehay LE, Ferguson DC, Ghoreishi-Haack NS, Grimm DR, Guan R, Han EK, Holley-Shanks RR, Hristov B, Idler KB, Jarvis K, Johnson EF, Kleinberg LR, Klinghofer V, Lasko LM, Liu X, Marsh KC, McGonigal TP, Meulbroek JA, Olson AM, Palma JP, Rodriguez LE, Shi Y, Stavropoulos JA, Tsurutani AC, Zhu GD, Rosenberg SH, Giranda VL, Frost DJ (2007) ABT-888, an orally active poly(ADP-ribose) polymerase inhibitor that potentiates DNA-damaging agents in preclinical tumor models. *Clin Cancer Res* 13(9):2728–2737
  35. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ, Ashworth A, Carmichael J, Kaye SB, Schellens JH, de Bono JS (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 361(2):123–134
  36. Troester MA, Hoadley KA, Sorlie T, Herbert BS, Borresen-Dale AL, Lonning PE, Shay JW, Kaufmann WK, Perou CM (2004) Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res* 64:4218–4226
  37. Gonzalez MA, Tachibana KE, Chin SF, Callagy G, Madine MA, Vowler SL, Pinder SE, Laskey RA, Coleman N (2004) Geminin predicts adverse clinical outcome in breast cancer by reflecting cell-cycle progression. *J Pathol* 204(2):121–130
  38. Graeser M, McCarthy A, Lord CJ, Savage K, Hills M, Salter J, Orr N, Parton M, Smith IE, Reis-Filho JS, Dowsett M, Ashworth A, Turner NC (2010) A marker of homologous recombination predicts pathologic complete response to neoadjuvant chemotherapy in primary breast cancer. *Clin Cancer Res* 16(24):6159–6168
  39. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, Dairkee S, Tokuyasu T, Ljung BM, Jain AN, McLennan J, Ziegler J, Chin K, Devries S, Feiler H, Gray JW, Waldman F, Pinkel D, Albertson DG (2006) Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 6(1):96
  40. Arnes JB, Brunet JS, Stefansson I, Begin LR, Wong N, Chappuis PO, Akslen LA, Foulkes WD (2005) Placental cadherin and the

- basal epithelial phenotype of BRCA1-related breast cancer. *Clin Cancer Res* 11(11):4003–4011
41. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, Trudel M, Akslen LA (2003) Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst* 95(19):1482–1485
  42. Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S (2006) X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* 9(2):121–132
  43. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320(5874):362–365
  44. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8:R76
  45. Cressman VL, Backlund DC, Hicks EM, Gowen LC, Godfrey V, Koller BH (1999) Mammary tumor formation in p53- and BRCA1-deficient mice. *Cell Growth Differ* 10(1):1–10
  46. Jiang Z, Deng T, Jones R, Li H, Herschkowitz JI, Liu JC, Weigman VJ, Tsao MS, Lane TF, Perou CM, Zacksenhaus E (2010) Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status. *J Clin Invest* 120(9):3296–3309. doi:[10.1172/JCI41490](https://doi.org/10.1172/JCI41490)
  47. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314(5797):268–274
  48. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 91(2):355–358
  49. Quinn JE, Kennedy RD, Mullan PB, Gilmore PM, Carty M, Johnston PG, Harkin DP (2003) BRCA1 functions as a differential modulator of chemotherapy-induced apoptosis. *Cancer Res* 63(19):6221–6228
  50. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434(7035):917–921