Supplementary Text

Clinical sample data

A discovery data set comprising 142 fresh-frozen, macro-dissected primary invasive TNBCs were obtained from patients treated at Guy's and St Thomas' Hospitals, London, UK (referred to as 'KCL data'). A detailed description of the clinico-pathological features has been published previously (1) and is provided in Supplementary Table S1. DNA from 126 tumours was sent to Atlas Biolabs GmbH (Berlin, Germany) for SNP6.0 genome profiles (Affymetrix, Santa Clara, CA, USA), out of which 88 had matching Affymetrix GeneChip Human Exon 1.0ST array gene expression data. Processing of the raw expression data, including classification to the PAM50 subtypes has been reported previously (1). Raw data from the gene expression microarray and SNP6.0 array experiments are available for download under the accession numbers GSE40267 and E-MTAB-2626, respectively.

For the METABRIC data set (2), the triple-negative status of invasive breast carcinomas was based on reported ER and HER2 immunohistochemistry (IHC) as well as *ESR1*, *PGR* and *ERBB2* gene expression to give a cohort of 115 TNBCs (list of samples and clinical data is shown in Supplementary Table S1). Raw probe-level expression data generated by the Illumina HumanHT-12 v3 Expression BeadChip (Illumina, Santa Diego, CA, USA) and SNP6.0 microarray data were acquired from the European Genome-Phenome Archive (3) accession number EGAD00010000164. Gene expression microarray data were normalised with ComBat (4).

Copy number, expression and clinical data were retrieved for the TCGA breast and ovarian cancer cohorts from (5). Eighty TNBCs were selected based on ER, PR and HER2 IHC and blood-matched SNP6.0 data, of which 52 had processed gene-level Agilent 244K Custom Gene Expression data.

The PrECOG cohort comprised 80 patients enrolled as part of a single-arm, phase II study (NCT00813956) in which gemcitabine and carboplatin plus iniparib were administered prior to surgical intervention (6). Response to carboplatin was stratified on the basis of the residual cancer burden (RCB) index whereby patients with an RCB > 1 were considered treatment-unresponsive and those with an RCB of 0 or 1 deemed responders. Seventy-eight of these patients had their DNA profiled using Affymetrix SNP6.0, and 75 of 78 were evaluable with respect to treatment response. For the TCGA HGSC data set, 299 patients were selected having had a platinum agent as first-line treatment, possessed grade 3, stage III/IV tumours, and had interpretable drug response data. Platinum responders were defined as those patients who demonstrated complete or partial response to treatment with no progression, recurrence or metastasis within 6 months of initial treatment. Platinum non-responders were defined as those who were annotated with progressive or stable disease following treatment, or progression, recurrence or metastasis within 6 months.

DNA copy number analysis

Raw Genome-wide SNP6.0 microarray data were intensity-normalised with Nexus Copy Number 6.0 software (BioDiscovery, Hawthorne, CA, USA), using HapMap270 samples provided by Affymetrix (Affymetrix, Santa Clara, CA, USA) as the reference genome for

the KCL, METABRIC and cell line data sets. For the TCGA TNBC and HGSC data sets, SNP6.0 data from matched blood normal samples was used as reference. Normalisation was followed by application of the circular binary segmentation-based SNPRank algorithm (BioDiscovery, Hawthorne, CA, USA). Next, allele-specific copy number profiling was performed with Tumor Aberration Prediction Suite (TAPS) (7) or, in the case of TCGA HGSCs, ASCAT (8), using the R statistical language environment 2.14.1 (9).

For the KCL data set, 111 of 126 samples passed TAPS, yielding a gene expression-overlapping cohort of 77 TNBCs. For the METABRIC, PrECOG, TCGA TNBC and HGSC data, 97 of 115, 71 of 75, all 80, and 299 of 316 samples, respectively, were successfully called by TAPS or ASCAT. SNP6.0 data for all cell lines passed TAPS. Failure to determine allele-specific copy number with clinical data was due to inadequate tumour content, excessive heterogeneity or poor segmentation.

Chromothripsis-like features

Chromosomal arms harbouring chromothripsis-like features were ascertained using the operational definitions described in (10). In brief, for each chromosome, the number of breakpoints was ascertained and the distribution of distances between those breakpoints compared to the distribution expected if those breakpoints occurred randomly along the chromosome using a Kolmogorov-Smirnov test. For each chromosomal arm, where (i) $P < 0.05$ for the whole chromosome comparison; (ii) the number of breakpoints was greater on that arm than the number on the other arm

weighted by length; (iii) and the number of copy number states for that arm was 2, we called chromothripsis for that arm.

Scores of chromosomal instability scarring (SCINS)

A Sweave document to reproduce SCINS is supplied in a separate file.

Procedural description of SCINS

Segments of allele-specific copy number profiles were categorised into one of three non-overlapping types: AiCNAs, AbCNAs and CnLOH, forming the basis for SCINS (Supplementary Table S2). SCINS are calculated using the following steps:

1. The proportion of the genome consisting of AiCNA segments, save those segments that encompass a whole chromosome, is calculated.
2. The number of AiCNA segments greater than or equal to 8 Mbp in length but less than the length of a whole chromosome is counted.
3. The measure of AiCNA segments (S_{AiCNA}) is calculated by multiplying the proportion obtained in step 1) by the number of segments counted in step 2).
4. The proportion of the genome consisting of CnLOH segments is calculated.
5. The number of CnLOH segments greater than or equal to 4 Mbp in length, including those that span a whole chromosome, is counted.
6. The measure of CnLOH segments (S_{CnLOH}) is calculated by multiplying the proportion obtained in step 4) by the number of segments counted in step 5).
7. The measure of AbCNA segments (S_{AbCNA}) is calculated by counting the number of AbCNA segments greater than or equal to 8 Mbp in length.
8. The measure of all allelic imbalanced segments (S_{Ai}) is calculated by summing S_{AiCNA} and S_{CnLOH}.

Formal description of SCINS

1. S_{AiCNA}

An AiCNA segment is defined as follows,

$$\text{AiCNA: } (x, y \in \mathbb{Z}) \wedge (x > 0) \wedge \left(\frac{y}{x+y} \neq 0.5\right), \tag{1.1}$$

where x is the major copy number of the segment, y is the minor copy number of the segment, and each segment has length l_{AiCNA} with units in Mbp.

Therefore,

$$S_{\text{AiCNA}} = \frac{\sum L_{\text{AiCNA}}}{l_{\text{genome}}} \times \#L_{\text{AiCNA}}, \tag{1.2}$$

where L_{AiCNA}: $8\ \text{Mbp} \leq l_{\text{AiCNA}} < l_{\text{chromosome}}$.

2. S_{CnLOH}

A CnLOH segment is defined as follows,

$$\mathrm{CnLOH}\colon (x = 2) \wedge (y = 0), \tag{2.1}$$

where x is the major copy number of the segment, y is the minor copy number of the segment, and each segment has length l_{CnLOH} with units in Mbp.

Therefore,

$$S_{\mathrm{CnLOH}} = \frac{\sum L_{\mathrm{CnLOH}}}{l_{\mathrm{genome}}} \times \#L_{\mathrm{CnLOH}}, \tag{2.2}$$

where $L_{\mathrm{CnLOH}}\colon l_{\mathrm{CnLOH}} \geq 4$ Mbp .

3. S_{AbCNA}

An AbCNA segment is defined as follows,

$$\text{AbCNA: } (x, y \in \mathbb{Z}) \wedge \left(\frac{y}{x+y} = 0.5\right), \tag{3.1}$$

where x is the major copy number of the segment, y is the minor copy number of the segment, and each segment has length l_{AbCNA} with units in Mbp.

Therefore,

$$S_{\text{AbCNA}} = \#L_{\text{AbCNA}}, \tag{3.2}$$

where $L_{\text{AbCNA}}: l_{\text{AbCNA}} \geq 8 \text{ Mbp}$.

4. S_{Ai}

$$S_{Ai} = S_{AiCNA} + S_{CnLOH} \quad (4.1)$$

Explanation of SCINS

Size thresholds for the segment count were adopted to (i) exclude the abundant short aberrations generated by solitary complex events such as chromothripsis; and (ii) exclude potential germline CNVs in the cases of S_{AiCNA} and S_{AbCNA}, and germline homozygous regions in the case of S_{CnLOH}. Since many of the breakpoints for long aberrations are enriched at known CNPs, a size threshold was considered more appropriate than CNP filtering (11). Moreover, use of a size threshold enabled us to draw a comparison between S_{CnLOH} – for which no filtering reference set equivalent to the Database of Genomic Variants was available – and the two CNA-based scores, S_{AiCNA} and S_{AbCNA}. The 8 Mbp threshold was chosen for S_{AiCNA} and S_{AbCNA} on the basis of the size of the largest documented CNP in the Database of Genomic Variants (12). Likewise, the 4 Mbp threshold was chosen to reflect the common observation of germline homozygous regions of up to 4 Mbp in outbred individuals (13) . To accommodate the possibility of multiple, separate events generating numerous, short segments, the counts of AiCNA and CnLOH segments were weighted by the fraction of the genome with AiCNA and CnLOH, respectively. S_{AbCNA} was developed to represent whole genome and chromosomal alterations, which are by definition large-scale alterations. As such, the number of AbCNA segments was not weighted by proportion.

Gene expression analysis

SAM was used to assess differential gene expression between SCINS-defined clusters using the samr R package (14). A two-class test with 1,000 permutations was employed

to compare expression between the combined Hi-AiCNA/Hi-CnLOH cluster, and the Lo-SCINS cluster. A multi-class test with 1,000 permutations was used to compare expression between the Hi-AiCNA, Hi-CnLOH and Lo-SCINS clusters.

For the meta-analysis approach, each gene was first assigned to the SCINS-defined cluster for which its mean expression was highest. For those genes that were assigned to the same SCINS-defined cluster in both the KCL and METABRIC data sets, the Stouffer z method was used to combine the q-values from each data set.

For gene signature analysis, 11 gene sets were retrieved from the literature (Supplementary Table S4) and the activation score of each signature was determined and evaluated as described previously (15). An expectation-maximisation algorithm was used to model the distribution of the bimodal *HORMAD1* expression in TNBCs, as a mixture of two normal distributions, thereby determining high and low *HORMAD1*-expressing tumours.

Meiotic genes

Genes documented to be involved in meiosis were identified by reference to the Reactome Pathway Database version 48 (http://www.reactome.org/).

Platinum response prediction in the PrECOG data set

ROC analyses were performed using S_{AiCNA}, S_{CnLOH}, S_{Ai}, *BRCA1/2* mutation status and dichotomised *HORMAD1* expression (established from its bimodal distribution) as predictors of platinum-based chemotherapeutic outcome (Supplementary Figure S13).

In addition, a combined *HORMAD1-BRCA1/2* marker was evaluated by considering the four possible combinations of these two markers: *HORMAD1* low-*BRCA1/2* wildtype; (ii) *HORMAD1* high-*BRCA1/2* wildtype; (iii) *HORMAD1* low-*BRCA1/2* mutant; and (iv) *HORMAD1* high-*BRCA1/2* mutant. Combined scar-*BRCA1/2* mutation markers were obtained by multiplying the scar score by '2' if *BRCA1/2* was mutated, and by '1' otherwise. Optimal cut-offs for the continuous and *HORMAD1-BRCA1/2* markers were established by finding the threshold that produced the highest balanced accuracy, which was defined as (sensitivity + specificity)/2. Univariable logistic regression for dichotomised *HORMAD1* expression was also performed and significance was assessed through a Wald test.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. Frequency of genomic aberrations in the KCL TNBC cohort. A, copy number gains (green; copy number > 2) and losses (red; copy number < 2) in 111 TNBCs from the KCL cohort in which the proportion of tumours with gain or loss (vertical axis) at each genomic location (x-axis) is plotted. **B**, the frequency of losses of heterozygosity defined as a minor allele copy number of 2 at a presumed heterozygous locus. **C**, the frequency of copy number neutral losses of heterozygosity defined as a minor allele copy number of 0 and a total copy number of 2.

Supplementary Figure S2. Conceptual diagrams of Scores of Chromosomal Instability Scarring (SCINS). The maternal chromosome of a pair of homologous chromosomes is in beige and the paternal chromosome in purple. Allelic regions of the genome are illustrated by lettered blocks of genome. Top left is an example of an allelic imbalance associated with a copy number gain, and to the right, an allelic imbalance associated with copy number loss. These are among the types of scar captured as part of the S_{AiCNA} measure, and may be generated through either NHEJ or non-conservative, non-allelic homologous recombination. Allelic-balanced copy number gains (bottom left) are an example of the types of scar captured by the S_{AbCNA} measure, many of which may be generated as a consequence of whole genome doubling. Copy number neutral losses of heterozygosity may be generated through allelic homologous recombination, and contribute to the S_{CnLOH} measure.

Supplementary Figure S3. Extent and location of SCINS in TNBC. A-C, bar plots showing the contributions for each sample in the METABRIC TNBC data set (n = 97) (**A**), TCGA TNBCs (n = 80) (**B**) and PrECOG TNBCs (n =71) (**C**) of S_{AbCNA} (yellow), S_{AiCNA} (red) and S_{CnLOH} (blue), as a fraction of the sample with the highest aggregate of these measures (right-most bar). Samples are displayed from left to right in ascending order. Each column in the heatmap underneath corresponds to a sample in the bar plot, and depicts the genomic location of each type of scar. Chromosome number on the vertical axis runs from 1 at the top to X at the bottom.

Supplementary Figure S4. Chromosome-wise distribution of SCINS across four TNBC cohorts. Barplots show for S_{AbCNA} (top, yellow), S_{AiCNA} (middle, red) and S_{CnLOH} (bottom, blue), the number of scar segments over the total number of segments (y-axis) on a given chromosome (x-axis) for **A**, KCL, **B**, METABRIC, **C**, PrECOG and **D**, TCGA TNBC datasets. Percentages at the top of each bar represent the fraction contributed by the y-axis metric for a given chromosome to the whole genome. The mean and standard deviation (SD) of the y-axis metrics for a dataset are shown at the top of each barplot.

Supplementary Figure S5. SCINS-defined tumour classes identify HORMAD1 to be among the top genes linked with Hi-AiCNA/Hi-CnLOH in PrECOG TNBCs. A, heatmap (yellow-red image) and dendrogram showing the results of clustering 59

TNBCs from the PrECOG dataset according to S_{AbCNA}, S_{AiCNA} and S_{CnLOH}. The three TNBC clusters identified include Lo-SCINS (green arm of dendrogram), Hi-AiCNA (red arm of dendogram) and Hi-CnLOH (blue arm of dendrogram). Scores were standardised by row mean and a colour scale is shown to the left. Labels for each SCINS measure are displayed on the left. Boxplots underneath depict the distribution of S_{AiCNA} and S_{CnLOH} across the three clusters, identifying for each dataset a Lo-SCINS (green arm of dendrogram), a Hi-AiCNA (red arm of dendogram) and a Hi-CnLOH (blue arm of dendrogram) group. **B**, volcano plot of the $\log_2$ fold change of expression for all genes against the SAM D-score. Positive D-scores indicate an association with the composite Hi-AiCNA/Hi-CnLOH cluster, whereas negative D-scores represent an association with the Lo-SCINS cluster. No genes were significantly associated with either cluster (i.e. FDR $Q < 0.1$; SAM) and so are all coloured in grey. *HORMAD1* is indicated in blue. **C**, the median expression across samples for each dataset of each gene that was assayed in both the KCL and METABRIC TNBC datasets was used to conduct gene-level correlation analysis between the two cohorts. The median relative expression for METABRIC is shown on the y-axis and that of the KCL TNBC cohort on the x-axis. Each point represents a gene. *HORMAD1* is denoted by a red cross. Spearman's rank correlation coefficient and associated p-value are shown.

Supplementary Figure S6. Gene expression signatures and *HORMAD1* expression. A-J, box plots of the weighted, median-centred expression of genes (vertical axes) present in previously published signatures of different aspects of genomic instability (16-21)(see Supplementary Table S4). HORMAD1 groups were

defined as described in the Materials and Methods, and SCINS cluster membership is denoted by the colour of the data points: red = Hi-AiCNA, blue = Hi-CnLOH, green = Lo-SCINS.

Supplementary Figure S7. Immunofluorescent localisation of HORMAD1 in cell lines. A, SUM159 transfected with a C-terminally GFP tagged HORMAD1 construct. **B**, staining of endogenous HORMAD1 in HCC1143 cells before and after induction of DNA damage by IR and HU. For IR cells were irradiated with 5gy and fixed 1 hour later. For HU, cells were treated for 16 h prior to fixation.

Supplementary Figure S8. Expression of HORMAD1 in breast cancer cell lines and tumour samples. A, Immunohistocytochemistry of HORMAD1 in two breast cancer cell lines (right) and in a primary TNBC (left). **B,** western blots of HORMAD1 in 4 cell lines and 19 primary TNBC, top. Actin normalized quantification of HORMAD1 protein levels are plotted for each sample. The gene expression levels as determined by the Affymetrix GeneChip Human Exon 1.0ST array are shown underneath each tumour sample.

Supplementary Figure S9. HORMAD1 inhibits activity of the DR-GFP HR reporter. A, SUM159-LacZ-V5 and SUM159-HORMAD1-V5 cells were seeded at equal densities and cell number measured after 6 days. Bars represent mean and error bars represent

SEM. **B,** DR-GFP reporter activity in CAL51 after HORMAD1 expression. BRCA2 siRNA is included as a positive control. Activity is normalized to CAL51 treated with non-targeting siRNA. Data represents the mean of triplicate wells of a representative experiment, error bars are the SEM. Absolute percentage of GFP positive cells is in green above bars. **C,** DR-GFP reporter activity in HCC1143 after HORMAD1 knockdown. BRCA2 siRNA is included as a positive control. Activity is normalized to HCC1143 treated with non-targeting siRNA. Data represents the mean of triplicate wells of a representative experiment, error bars are the SEM. Absolute percentage of GFP positive cells is in green above bars. **D,** mouse ES cells with a chromosomally integrated DR-GFP reporter were transfected with pCAGGS empty vector, pEZ-M67 HORMAD1, or pCAGGS HORMAD1 and HR activity assessed by quantifying GFP positive cells. Bars represent the mean and SEM of 5 independent experiments. Mean absolute percentage of GFP positive cells is in green above bars. Western blot shows expression of HORMAD1 by both vectors. E**,** basal γ-H2AX foci in LacZ and HORMAD1 overexpressing SUM159 (green = γ-H2AX*,* blue = nuclei)**. F**, RAD51 foci in irradiated and HU treated cells in LacZ and HORMAD1 overexpressing SUM159 (red = RAD51, blue = nuclei).

Supplementary Figure S10. Effect of HORMAD1 overexpression and knockdown on cell growth and proliferation. A, Effect of HORMAD1 transfection on cell growth of SUM159 cells. Cells were seeded at equal densities, transfected with HORMAD1 or empty vector and cell number measured after 6 days. **B**, Effect of HORMAD1 transfection on cell cycle distribution of SUM159 cells. SUM159 cells were transfected

with empty or HORMAD1 expression vector and cell cycle distribution assessed by FACS after 72 hours **C,** effect of HORMAD1 knockdown on cell cycle distribution in HCC1143, left, and MDAMB436, right.

Supplementary Figure S11. HORMAD1 promotes activity of the EJ5 NHEJ reporter. A, 53BP1 foci in irradiated and HU treated cells in lacZ and HORMAD1 overexpressing SUM159 (green = 53BP1, blue = nuclei). **B,** EJ5 NHEJ reporter activity in SUM159 after HORMAD1 expression. Activity is normalized to SUM159 transfected with empty vector. Left, representative FACS plots showing GFP positive cells. Right, data represents the mean of triplicate wells of a representative experiment, error bars are the SEM. **C,** the EJ5 NHEJ reporter vector was used to assess NHEJ activity in HORMAD1 transfected SUM159. NHEJ activity is presented as a percentage compared to control empty vector. Bars represent the mean of 3 independent transfections. Absolute percentage of GFP positive cells is in green above bars.

Supplementary Figure S12. Effect of HORMAD1 knockdown on 53BP1 and RAD51 foci in HCC1143. A, IR induced 53BP1 focus formation in SUM159-LacZ-V5 or SUM159-HORMAD1-V5 cells. Data points represent the number of 53BP1 foci in individual nuclei and bars represent mean and standard deviation for each condition. Statistical significance of the mean number of foci between each group was assessed by Student's t-test**;** left, western blot showing HORMAD1 knockdown produced by HORMAD1 shRNA in HCC1143. **B**, IR induced RAD51 focus formation in SUM159-

LacZ-V5 or SUM159-HORMAD1-V5 cells. Data points represent the number of RAD51 foci in individual nuclei and bars represent mean and standard deviation for each condition. Statistical significance of the mean number of foci between each group was assessed by Student's t-test.

Supplementary Figure S13. *BRCA1/2* mutation, platinum sensitivity and genomic scarring or *HORMAD1* expression. Box plots depicting the distribution in TCGA HGSCs of S_{AbCNA} (**A**), N_{tAi} and S_{LOH} (**B**) and *HORMAD1* expression (**C**), and then in PrECOG TNBCs, of S_{AbCNA} (**D**), N_{tAi} and S_{LOH} (**E**) and *HORMAD1* expression (**F**, left panel). Tumours have been stratified by platinum agent responder status, and sub-stratified by *BRCA1*/*2* mutation status. P-values were obtained using a Wilcoxon rank-sum test.

Supplementary Figure S14. Effect of HORMAD1 knockdown on drug sensitivity in HCC1143. HCC1143 cells were transfected with non-targeting (NT) or HORMAD1 siRNA and plated in 6 well plates and then exposed to various doses of **A**, BMN673, or **B**, cisplatin for 7 days. Statistical significance was assessed using ANOVA. Error bars represent the standard error of the mean from triplicate wells.

Supplementary Figure S15. ROC analysis of PrECOG TNBCs. S_{AiCNA} (red curve), S_{CnLOH} (blue curve), S_{Ai} (grey curve), *BRCA1/2* mutation status (purple point) and

dichotomised *HORMAD1* expression (green point) were used as predictors of platinum-based chemotherapeutic response in a ROC analysis. Sensitivity and 1-Specificity are shown on the vertical and horizontal axes, respectively.

Supplementary Figure S16. *HORMAD1* expression in unselected and *BRCA1/2* wildtype TNBCs. Bimodal distributions of *HORMAD1* expression in **A**, the whole PrECOG dataset and **B**, BRCA1/2 wildtype PrECOG tumours. Light and dark grey curves depict the distribution of low and high *HORMAD1*-expressing tumours, respectively. Maximum Likelihood Estimation was used to define the cut-off between high and low *HORMAD1* expressing tumours, which are very similar between the two.

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table S1. KCL TNBC and breast cell line cohort characteristics. Sheet 1, showing the characteristics for the 111 TNBCs used in this study. Sheet 2, showing the names and characteristics of the 38 breast cell lines used.

Supplementary Table S2. Categories of genomic segment used as the basis for measuring SCINS. The genomic profiles derived from SNP6.0 Affymetrix Microarrays of TNBCs, HGSCs and cell lines were classified into one of AiCNA (allelic-imbalanced CNAs), AbCNA (allelic-balanced CNAs) and CnLOH (copy neutral LOH), or *No scar* for

which *x* denotes the integer copy number of the major allele of a given genomic segment; and *y* denotes the integer copy number of the minor allele of the segment.

Supplementary Table S3. Genes expressed in a SAM of different SCINS clusters. Genes expressed in a SAM of different SCINS clusters. Sheets 1 and 2, showing genes with a Q < 0.1, their different mean expressions among the three SCINS clusters (Hi-AiCNA, Hi-CnLOH and Lo-SCINS), and multi-class SAM q-value and D-scores for the KCL (Sheet 1) and METABRIC (Sheet 2) data sets. Genes are ordered by D-score. For the "Concordant" column, a '1' indicates that the gene appeared in both datasets with the same SCINS-defined cluster assignment while a '0' indicates otherwise. Sheets 3 and 4, the genes with a Q < 0.1 in a two-class SAM between the Lo-SCINS cluster and the combined Hi-AiCNA/Hi-CnLOH clusters in the KCL (Sheet 3) and METABRIC (Sheet 4) data sets. Genes are ordered by fold change, by which *HORMAD1* lies at the top for the KCL data set and places sixth in METABRIC.

Supplementary Table S4. Gene expression signatures of genomic instability. Tables showing the literature references and compositions of 11 different genomic instability-related gene expression signatures used to explore the relationships to SCINS and *HORMAD1* expression.

Reference:

1. Gazinska P, Grigoriadis A, Brown JP, Millis RR, Mera A, Gillett CE, et al. Comparison of basal-like triple-negative breast cancer defined by morphology, immunohistochemistry and transcriptional profiles. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2013;26:955-66.

2. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012.

3. EGA. Available from: http://www.ega.ac.uk

4. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118-27.

5. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61-70.

6. Telli ML, Jensen KC, Abkevich V, Hartman AR, Vinayak S, Lanchbury J, et al. Homologous Recombination Deficiency (HRD) score predicts pathologic response following neoadjuvant platinum-based therapy in triple-negative and BRCA1/2 mutation-associated breast cancer (BC) San Antonio Breast Cancer Symposium; 2012 December 15, 2012; San Antonio. p. PD09-4.

7. Rasmussen M, Sundstrom M, Goransson Kultima H, Botling J, Micke P, Birgisson H, et al. Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. Genome biology. 2011;12:R108.

8. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:16910-5.

9. 2011 RDCT. Available from: http://www.r-project.org/

10. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011;144:27-40.

11. Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer discovery. 2012;2:366-75.

12. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nature genetics. 2004;36:949-51.

13. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-lauc L, et al. Runs of Homozygosity in European Populations. The American Journal of Human Genetics. 2008;83:658.

14. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:5116-21.

15. Grigoriadis A, Mackay A, Noel E, Wu PJ, Natrajan R, Frankum J, et al. Molecular characterisation of cell line models for triple-negative breast cancers. BMC Genomics. 2012;13:619.

16. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nature genetics. 2006;38:1043-8.

17. Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, et al. A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. PloS one. 2009;4:e6693.

18. Cheng WY, Ou Yang TH, Anastassiou D. Development of a prognostic model for breast cancer survival in an open challenge environment. Science translational medicine. 2013;5:181ra50.

19. Coutant C, Rouzier R, Qi Y, Lehmann-Che J, Bianchini G, Iwamoto T, et al. Distinct p53 gene signatures are needed to predict prognosis and response to chemotherapy in ER-positive and ER-negative breast cancers. Clinical cancer research : an official journal of the American Association for Cancer Research. 2011;17:2591-601.

20. Ertel A, Dean JL, Rui H, Liu C, Witkiewicz AK, Knudsen KE, et al. RB-pathway disruption in breast cancer: differential association with disease subtypes, disease-specific prognosis and therapeutic response. Cell cycle. 2010;9:4153-63.

21. Loi S, Haibe-Kains B, Majjaj S, Lallemand F, Durbecq V, Larsimont D, et al. PIK3CA mutations associated with gene signature of low mTORC1 signaling and better

outcomes in estrogen receptor-positive breast cancer. Proceedings of the National Academy of Sciences of the United States of America. 2010;107:10208-13.