MOLECULAR ONCOLOGY 10 (2016) 179–193

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

Personalised pathway analysis reveals association between DNA repair pathway dysregulation and chromosomal instability in sporadic breast cancer

Chao Liu[a], *Sriganesh Srihari*[a], *Samir Lal*[b], *Benoît Gautier*[c], *Peter T. Simpson*[b,d], *Kum Kum Khanna*[e], *Mark A. Ragan*[a,**], *Kim-Anh Lê Cao*[c,*]

[a]*Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4067, Australia*
[b]*The University of Queensland, UQ Centre for Clinical Research, Herston, QLD 4029, Australia*
[c]*University of Queensland Diamantina Institute, Translational Research Institute, Woolloongabba, QLD 4102, Australia*
[d]*School of Medicine, The University of Queensland, Herston, QLD 4006, Australia*
[e]*QIMR-Berghofer Medical Research Institute, Herston, Brisbane, QLD 4029, Australia*

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form
19 August 2015
Accepted 4 September 2015
Available online 26 September 2015

Keywords:
DNA repair
Homologous recombination
Breast cancer
Chromosomal instability
Pathway analysis

ABSTRACT

The Homologous Recombination (HR) pathway is crucial for the repair of DNA double-strand breaks (DSBs) generated during DNA replication. Defects in HR repair have been linked to the initiation and development of a wide variety of human malignancies, and exploited in chemical, radiological and targeted therapies. In this study, we performed a personalised pathway analysis independently for four large sporadic breast cancer cohorts to investigate the status of HR pathway dysregulation in individual sporadic breast tumours, its association with HR repair deficiency and its impact on tumour characteristics. Specifically, we first manually curated a list of HR genes according to our recent review on this pathway (Liu et al., 2014), and then applied a personalised pathway analysis method named Pathifier (Drier et al., 2013) on the expression levels of the curated genes to obtain an *HR* score quantifying HR pathway dysregulation in individual tumours. Based on the score, we observed a great diversity in HR dysregulation between and within gene expression-based breast cancer subtypes, and by using two published HR-defect signatures, we found HR pathway dysregulation reflects HR repair deficiency. Furthermore, we identified a novel association between HR pathway dysregulation and chromosomal instability (CIN) in sporadic breast cancer. Although CIN has long been considered as a hallmark of most solid tumours, with recent extensive studies highlighting its importance in tumour evolution and drug resistance, the molecular basis of CIN in sporadic cancers remains poorly understood. Our results imply that HR pathway dysregulation might contribute to CIN in sporadic breast cancer.

* *Corresponding author*. University of Queensland Diamantina Institute, Translational Research Institute, Level 5, 37 Kent Street, Woolloongabba, QLD 4102, Australia. Tel.: +61 7 3443 7069; fax: +61 7 3443 6966.
** *Corresponding author*. Institute for Molecular Bioscience, The University of Queensland, Level 6, 306 Carmody Rd, St. Lucia, QLD 4067, Australia. Tel.: +61 7 3346 2616; fax: +61 7 3346 2101.
E-mail addresses: m.ragan@uq.edu.au (M.A. Ragan), k.lecao@uq.edu.au (K.-A. Lê Cao).
http://dx.doi.org/10.1016/j.molonc.2015.09.007

1. Introduction

Chromosomal instability (CIN), defined as an increased rate of gain or loss of whole chromosomes or large chromosomal fragments, is a hallmark of most solid tumours. CIN is the primary form of genomic instability that is thought to be the major cause of genetic heterogeneity in cancer (Burrell et al., 2013b), and is thus strongly implicated in tumour evolution. CIN also has important clinical implications, as it has been linked to poor prognosis e.g. by conferring intrinsic multidrug resistance (Lee et al., 2011). The molecular basis of CIN in hereditary cancer is relatively clear, and has been attributed to mutations in DNA repair genes (Negrini et al., 2010); however, the underlying mechanisms of CIN in various sporadic cancers remain poorly understood. Carter and colleagues developed a gene expression-based CIN signature, termed CIN25, based on 25 genes that are most overexpressed in tumours with CIN (Carter et al., 2006). A considerable number of genes involved in *replication* and *cell cycle* contribute to this signature, suggesting an important link between these cellular processes and CIN. This was further corroborated by Negrini et al. (2010), who proposed a *replication stress* model to explain CIN in sporadic tumours; this model was recently validated in colorectal cancer (Burrell et al., 2013a).

Highly proliferative cancer cells undergo considerable replication stress that results in the stalling of replication forks. These stalled forks are usually stabilised and restarted after the source of stress is removed via a complex replication stress response pathway (Zeman and Cimprich, 2014). Lack of stabilisation and/or the prolonged persistence of a stalled fork can generate DNA double-strand breaks (DSBs), which are subsequently repaired by DSB repair machinery to restart the forks. However, in the absence of such a DSB repair machinery the DSBs will develop into chromosomal breaks, resulting in CIN. *Homologous recombination* (HR) is a crucial pathway responsible for repairing DSBs during replication. Using homologous sister chromatid as templates, HR presents a high-fidelity repair mechanism that is crucial for error-free DNA replication.

The core components of HR are fairly well established for their specific roles i.e. monitoring, signalling and repairing of DSBs (Liu et al., 2014), and HR defects can be detected by investigating the loss-of-function mutations in these genes. However, the dysfunction of HR can also be caused by numerous other mechanisms. For example, changes or defects in chromatin remodelling (Price and D'Andrea, 2013; van Attikum and Gasser, 2009), microRNAs (Chowdhury et al., 2013; d'Adda di Fagagna, 2014; Sharma and Misteli, 2013), post-translational modifications such as ubiquitination and sumoylation (Bekker-Jensen and Mailand, 2011; Dou et al., 2011; Ulrich, 2012), and inappropriate expression of certain genes that are not directly involved in HR (Y. Peng et al., 2015; Watkins et al., 2015) can considerably affect HR components, thereby causing aberrant HR function. As a consequence, single-gene approaches or approaches focussing on one mechanism yield only an incomplete picture of abnormal HR in a given tumour. On the other hand, HR-deficient cells may compensate for the defect in a given HR gene by altering the expression level of other HR genes (Pitroda et al., 2014). The most notable example is the overexpression of DNA repair protein RAD51 homolog 1 (*RAD51*), which is observed when breast cancer susceptibility gene 1 (*BRCA1*) (Martin et al., 2007), breast cancer susceptibility gene 2 (*BRCA2*) (Brown and Holt, 2009) or other key HR genes (Takata et al., 2001) are defective. It is therefore of interest to determine a measure of HR *pathway* dysregulation, aggregating the expression of all HR genes, which may reflect HR repair deficiency in tumours regardless of the mechanism that has led to the deficiency.

The vast majority of breast tumours are sporadic, accounting for 90%–95% of all diagnosed breast cancer cases (Davis, 2011) and are characterised by their great heterogeneity in biological property and patient outcome. To dissect this heterogeneity, estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) have been used as standardised diagnostic markers in clinical practice to guide the choice of treatment. Gene-expression profiling has defined five intrinsic subtypes (also known as PAM50 subtypes) with clinical relevance: Luminal A, Luminal B, Basal-like, HER2 and Normal-like (Hu et al., 2006; Parker et al., 2009; Perou et al., 2000; Sørlie et al., 2001). More-recent genomic studies, notably from the Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), have uncovered substantial heterogeneities within these receptor- or gene expression-based subtypes, resulting in the definition of up to ten subtypes (Ciriello et al., 2013; Curtis et al., 2012; Koboldt et al., 2012; Lehmann et al., 2011; Yanagawa et al., 2012). However, it is likely that heterogeneity exists even within these newly established subtypes. In the coming age of personalised medicine, each tumour may be analysed individually.

Pathway analysis has become the first choice to gain functional insights from expression data, beyond the detection of differential genes. Numerous pathway analysis tools have been developed; however, most of them are designed for providing pathway dysregulation information at population level instead of tumour level. Among the recently proposed methods for personalised pathways analysis (Ahn et al., 2014; Drier et al., 2013; Vaske et al., 2010; Wang et al., 2015a,b), Pathifier (Drier et al., 2013) has proven to be particularly robust. It has been successfully applied to provide a pathway-based classification of breast cancer (Livshits et al., 2015), and when combined with Cox regression and L1 penalised estimation, has achieved better prognosis prediction compared with gene-based models (Huang et al., 2014).

In this study, we sought to perform a personalised pathway analysis to obtain a comprehensive understanding of the status of HR pathway dysregulation in individual sporadic breast tumours, its association with HR repair deficiency and its impact on tumour characteristics (CIN in this case). To this end, we calculated for each breast tumour an *HR* score that quantified the extent of HR pathway dysregulation in that tumour. Based on the score, we observed a great diversity in HR dysregulation between and within the PAM50 subtypes, and by using two published HR-defect signatures, we found HR pathway dysregulation reflects HR repair deficiency. More importantly, we uncovered a novel association between HR dysregulation and CIN, which indicates that dysregulated

HR might contribute to replication stress-induced CIN in breast cancer. This knowledge may help future studies to identify the causative factors of CIN in sporadic breast cancer as well as in other cancer types.

2. Materials and methods

2.1. Genomic data

Whole-genome gene-expression data, DNA copy-number data, gene mutation data (only available for the TCGA samples) and related clinical data for four breast cancer cohorts (Table 1) were obtained from METABRIC (Curtis et al., 2012) and TCGA (Koboldt et al., 2012).

Gene-expression data and chromosomal-level DNA copy-number data from the METABRIC project (Genome-phenome Archive accession number EGAS00000000083) were made available upon request, and had already been preprocessed as described by Curtis et al. (Curtis et al., 2012). Gene-expression data from this project were based on the Illumina HT-12 v3 Expression Beadchip (Illumina, San Diego, CA, USA). The probe-level transcription estimates were mapped to gene-level estimates using the HT-12 v3 annotation file downloaded from the Illumina website (http://www.illumina.com/). Where two or more probes represented the same gene, the probe with the largest variation was chosen as the gene representative. DNA copy-number data from METABRIC had been generated using Affymetrix SNP 6.0 arrays (Affymetrix, Santa Clara, CA, USA). The corresponding PAM50 subtype assignment and clinical outcome were obtained from (Curtis et al., 2012).

The preprocessed gene-expression and DNA copy-number data (both chromosome-level and gene-level) for the TCGA RNA-seq cohort were downloaded via the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu/) on 13 October 2014. Gene-expression data for this cohort were measured using the Illumina HiSeq 2000 RNA Sequencing platform, and show the Expectation Maximization (RSEM)-normalised and percentile-ranked gene-level transcription estimates. DNA copy-number data for this cohort had been generated using Affymetrix SNA 6.0 arrays, with germline copy-number variation filtered out. PAM50 classifications for this cohort were obtained through personal communication with the TCGA consortium. A subset of these 1068 cases also has gene-expression data obtained from microarray. The Level 3 gene-expression data for this TCGA Microarray cohort and the corresponding PAM50 classifications were downloaded from the TCGA data portal publication site (https://tcga-data.nci.nih.gov/docs/publications/brca_2012/) on 3 June 2014. These gene-expression data were based on Agilent custom 244K whole-genome microarrays and had been preprocessed as described by Koboldt et al. (Koboldt et al., 2012). DNA copy-number data for this cohort were obtained as a subset of the TCGA RNA-seq cohort, as the samples of the former cohort were covered by the later cohort.

The preprocessed gene mutation data for 982 TCGA samples, generated on an Illumina GA system, were downloaded via the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu/) on 6 July 2015. Each gene had been assigned a value of 1 or 0, indicating whether a non-silent mutation was identified in the coding region of that gene (value = 1) or not (value = 0). These data were matched to the two TCGA cohorts respectively according to the sample ID.

2.2. HR pathway curation and calculation of HR score

Based on our recent review of the HR pathway (Liu et al., 2014), we manually curated a list of 82 genes with direct relevance to HR (Supplementary Table S1). We then applied Pathifier (Drier et al., 2013) to the mRNA expression level of the curated HR genes to calculate an *HR* score that quantifies HR pathway dysregulation in individual breast tumours. Based on gene-expression profiles for tumours and normal breast tissues, Pathifier transforms HR gene-expression measurements into a measure of HR pathway dysregulation by fitting a principal curve (see Supplementary Figure S1 for a visualisation of the curve) that captures the maximal variability of the expression levels of the HR genes in all samples, and then projects each sample onto that curve. A sample's *HR* score is defined as its distance along the curve from the centroid of the normal tissues (Drier et al., 2013).

Not all HR genes we curated were present in the gene-expression data for each of the four cohorts. We therefore calculated the *HR* score for each cohort based only on HR genes that are available for that cohort (ranges from 67 to 72, see Supplementary Table S1). No other ways for selecting HR genes were examined to minimize retrospective optimization for the correlations with CIN (see below).

2.3. CIN measurements calculation

The numbers of chromosomal breakpoints and the proportions of the genome affected by copy-number change (Genomic Instability Index, GII) for samples in the two METABRIC cohorts were downloaded from a recent study (Vollan et al., 2015) in which the METABRIC Group was involved.

Table 1 — Breast cancer cohorts analysed in this study.

Cohort	No. of tumour samples						No. of normal breast tissues
	All	Basal-like	HER2	LumA	LumB	Normal-like	
METABRIC Discovery	997	118	87	466	268	58	144
METABRIC Validation	995	213	153	255	224	144	144
TCGA RNA-seq	1068	188	80	549	213	38	113
TCGA Microarray	522	98	58	231	127	8	22

According to this study, a few samples with mismatched DNA/RNA were identified and excluded, resulting in 985 samples remaining in the Discovery cohort and 965 in the Validation cohort. To get the number of amplified/deleted genes for the same samples, we first calculated the copy number of each gene using the chromosomal-level DNA copy-number data available for the two cohorts, then applied cut-offs ($\geq$0.10 for amplified genes and ≤ -0.15 for deleted genes) that are similar to those used by METABRIC to define chromosomal regions with amplifications or deletions.

For the two TCGA cohorts, we used the chromosomal-level DNA copy-number data to calculate number of breaks by counting the total number of chromosomal segments at least 1 kb in length. The calculation of GII was also based on the chromosomal-level DNA copy-number data after filtering out segments shorter than 1 kb, and the same cut-offs as mentioned above ($\geq$0.10 for amplification and ≤ -0.15 for deletion) were used to identify chromosomal regions with copy-number change. The number of amplified/deleted genes for each of the two TCGA cohorts was obtained from the downloaded gene-level DNA copy-number data, where +1 and +2 represent amplification and −1 and −2 represent deletion.

2.4. *Survival analysis*

Survival analysis for both of the METABRIC datasets was performed using the R package *survival* (http://cran.r-project.org/web/packages/survival/index.html). Patient follow-up time was limited to 15 years, and only breast cancer-related deaths were counted.

3. Results

3.1. *An HR score for quantifying HR pathway dysregulation in individual breast tumours*

An *HR* score was developed for each breast tumour to quantify HR pathway dysregulation in that tumour; a high *HR* score means that the expression of the HR genes as a whole in an individual tumour is very different from the situation in normal breast tissues (see Supplementary Figure S2 for HR gene expression in tumours with low to high *HR* score). To calculate this score, we first manually curated a list of 82 HR genes (Supplementary Table S1) according to our recent review on the HR pathway (Liu et al., 2014). This gene list provides more up-to-date knowledge about the content of HR compared to publicly available pathway databases; for instance, it catalogues 54 more genes than the HR pathway in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000). The expression profiles of the curated HR genes were then employed as input to the Pathifier method (Drier et al., 2013) to compute the score. To ensure reproducibility of the results, we performed this pathway analysis independently for four large breast cancer cohorts that also include data on normal breast tissues (Table 1). Depending on data availability, the number of HR genes for calculating the score is slightly different across the cohorts (Supplementary Table S1).

The boxplots in Figure 1 display the *HR* score distribution in each cohort with regard to the PAM50 molecular subtypes, and in normal breast tissues. We observed a consistent pattern across the four cohorts: basal-like tumours generally have the highest *HR* score, followed by HER2 and Luminal B tumours, and then Luminal A and Normal-like tumours; the normal breast tissues always have the lowest *HR* score as a consequence of being the benchmark. Similar results can be seen in Supplementary Figure S3 showing *HR* score versus the *HR* score-based rank of the tumours of different subtypes. The consistent distribution of the *HR* score by tumour subtype across the different cohorts and gene-expression profiling platforms (RNA-seq and microarray in TCGA) is strong evidence that the *HR* score is robust and reproducible. Interestingly, we observed some variability in *HR* score within tumours of the same subtype, as highlighted by some outliers in the boxplots, suggesting some heterogeneity in HR pathway dysregulation within the subtypes.

3.2. *The HR score is reflective of HR repair deficiency*

The *HR* score is gene expression-based, and measures the extent to which the HR pathway is dysregulated. To test whether there exists an association between HR pathway dysregulation and HR repair deficiency, we next asked whether the *HR* score is reflective of HR repair deficiency (i.e., whether a tumour with high *HR* score is likely to be HR-defective). We used two published HR-defect signatures, Homologous Recombination Defect (HRD) (G. Peng et al., 2014) and Large-scale transitions (LSTs) (Popova et al., 2012), to test this hypothesis.

3.2.1. *Comparison with the HRD signature*

The HRD signature encompasses 230 genes that are differentially expressed between HR-intact and HR-deficient cells, and is intended to represent the global impact of HR defect on the transcriptome of a tumour cell (G. Peng et al., 2014). To identify tumours (or cell lines) with HR deficiency, Peng et al. performed a hierarchical clustering analysis based on the expression level of the 230 genes to divide samples into two clusters, one considered as HR-intact and the other HR-deficient (G. Peng et al., 2014).

In this study, we performed the same clustering analysis for each of the four cohorts (Figure 2A for the METABRIC Discovery cohort and Supplementary Figures S4–S6 for the three remaining cohorts). As shown in Figure 2A, tumours with low HR score (upper horizontal bar, green) are mostly tumours belonging to the HR-intact cluster, whereas tumours with high HR score (upper horizontal bar, red) are mostly tumours belonging to the HR-deficient cluster. To be more precise, Figure 2B shows the distribution of the HR score in the two HRD-based clusters for each of the four cohorts, demonstrating that tumours in the HR-deficient cluster in general have significantly higher HR score compared with tumours in the HR-intact cluster (p-values $\leq$9.1e-63, Wilcoxon rank-sum test). These observations indicate that tumours with high HR scores are likely to be HR-defective, as predicted by the HRD signature.

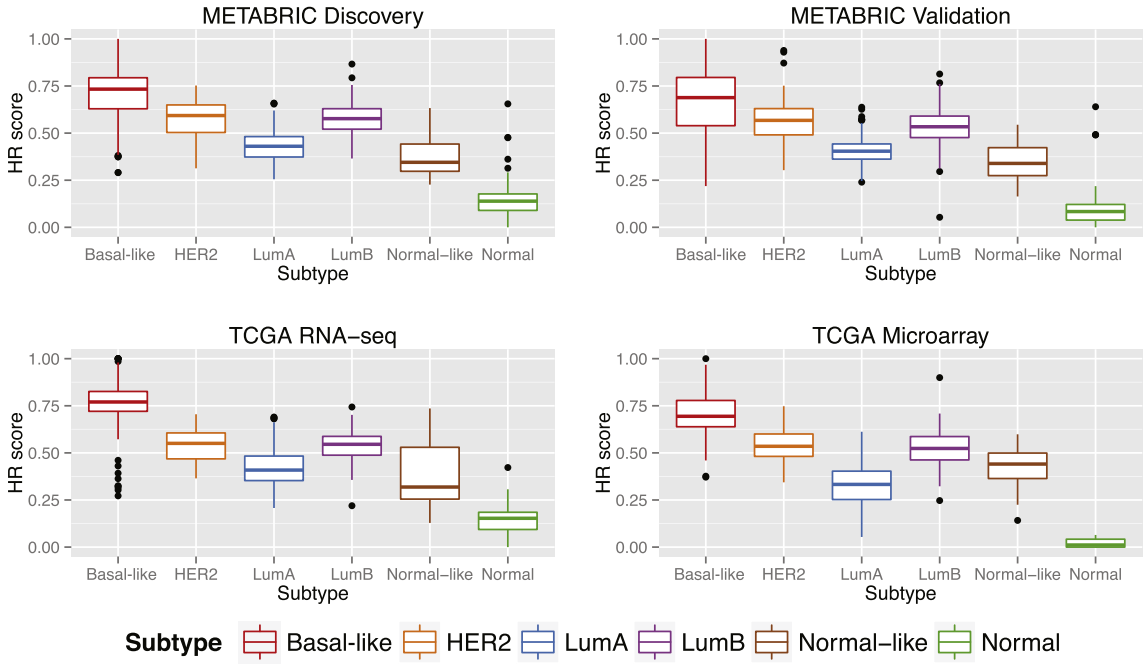

Figure 1 – Distribution of the *HR* score across the PAM50 subtypes and normal breast tissues (in green) for the four cohorts.

3.2.2. *Comparison with the LST signature*

LST refers to a chromosomal break whose flanking regions are at least 10 Mb in size. A tumour with a large number of LSTs indicates HR defect-related genomic scarring as a measure of chromosomal instability (Popova et al., 2012). In this study, we estimated the number of LSTs for each tumour using the DNA copy-number data, and divided each cohort into two groups according to the method and cut-offs described in (Popova et al., 2012): LST^{+} (≥20 LSTs) and LST^{-} (<20 LSTs). The numbers of LST^{+} and LST^{-} tumours identified in each cohort are summarised in Supplementary Table S2. As in the comparison with the HRD signature, we found that LST^{+} tumours generally have higher HR scores compared with LST^{-} tumours, even in the case of the METABRIC Discovery cohort where only nine LST^{+} tumours were identified (Figure 3). This observation also supports the idea that the *HR* score is indicative of HR defect.

Taken together, the results based on HRD and on LST demonstrate an association between HR pathway dysregulation, as represented by the *HR* score, and HR repair deficiency. In addition, in the two TCGA cohorts for which gene mutation data are available, we also observe that tumours with at least one non-synonymous mutation in one of six key HR genes have significantly higher *HR* score than do the tumours with no mutation in any of these genes (see Supplementary Figure S7 for more details). All these results support the existence of a compensatory mechanism through which HR-deficient cells respond to their HR defect by altering the expression level of HR genes. Interestingly, it has been proposed that melanoma cells exploit the overexpression of DNA repair genes, particular those involved in DSB repair, to increase their DNA repair capacity that is necessary for them to invade and give rise to distant metastases (Sarasin and Kauffmann, 2008). Consistent with this, overexpression of certain DNA repair genes is utilised by polyploid cells to overcome replication stress-induced senescence barriers (Zheng et al., 2012). All these results indicate that altering the expression of DNA repair genes or pathways may be a compensatory mechanism commonly exploited by tumour cells.

3.3. *Association with CIN*

Because replication stress has emerged as a common source of CIN in caner, and HR is the crucial pathway for the repair of replication stress-induced DSBs, we hypothesised that there might be a link between HR pathway dysregulation, which is indicative of HR repair deficiency as described above, and the degree of CIN in breast carcinomas. To test this hypothesis, we first examined the correlation between the *HR* score and the widely used CIN signature CIN25 (Carter et al., 2006). We then investigated the association between the *HR* score and each of the three common CIN measurements: number of chromosomal breakpoints, fraction of the genome with copy-number alterations (genomic instability index, GII), and number of amplified/deleted genes. In particular, as data pre-processing and segregation algorithms can significantly affect the actual value of the CIN measurements, we downloaded the numbers of chromosomal breaks and GII for the two METABRIC cohorts from a recent publication (Vollan et al., 2015). We believe these measures from a third-party study provide more-objective results for our analysis.

3.3.1. *Association with CIN25*

Figure 4 displays a scatter plot between the CIN25 score, defined as the mean expression value of the CIN25 genes (Carter et al., 2006), and the *HR* score for tumours from each of the four cohorts. Each cohort showed a high correlation

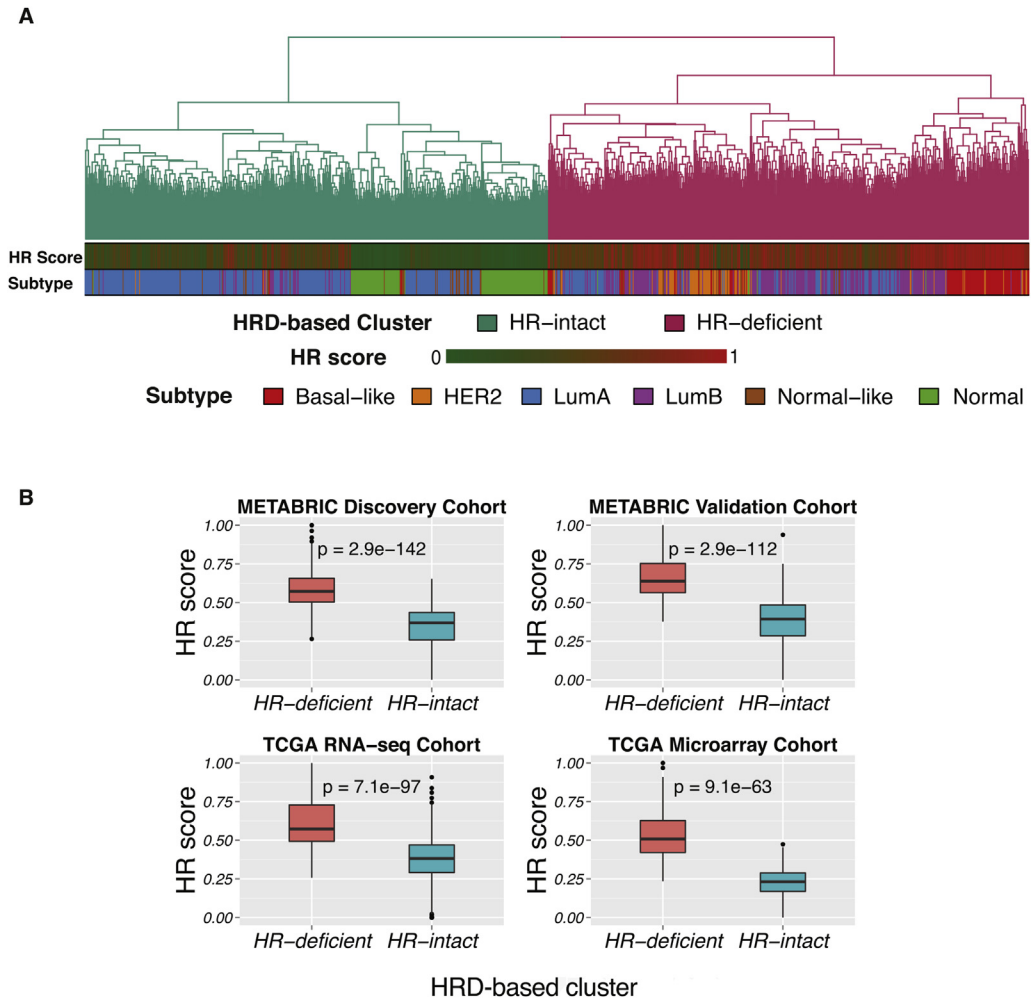

Figure 2 – Comparison of the *HR* score with the HRD signature. A: HRD-based hierarchical clustering of tumours from the METABRIC Discovery cohort. B: Distribution of the *HR* score in the two HRD-based clusters for each of the four cohorts. Colour represents the HRD-based cluster. The p-values were obtained using a Wilcoxon rank-sum test.

between the CIN25 score and the *HR* score (Spearman correlation coefficient r = 0.94 and r = 0.93 for the two METABRIC cohorts, and r = 0.85 and r = 0.96 for the two TCGA cohorts), indicating that the *HR* score is also correlated with CIN level. Moreover we found ten of the CIN25 genes (40%) to be present among the 230 genes of the HRD signature mentioned in Section 2.1, which indicates that HR defects might be one of the underlying biological mechanisms responsible for the expression change of the CIN25 genes.

Overall, these results revealed that the *HR* score correlates with the CIN25 score, and support the hypothesis that there exists an association between HR pathway dysregulation, as represented by the *HR* score, and CIN level in tumours, as predicted by the CIN25 score.

3.3.2. *Association with three common CIN measurements*

Because the CIN25 score only indirectly estimates CIN level in tumours, we also directly assessed the relationship between the *HR* score and each of the three common CIN measures (breakpoints, GII and number of amplified/deleted genes). We asked whether tumours with higher *HR* score tend to have a higher CIN level. To address this, we divided tumours into four equal-sized groups based on the *HR* score quartiles, and statistically examined the differences between adjacent groups for each of the three CIN measurements. The boxplots in Figure 5 (METABRIC Discovery cohort) show a high variability in each *HR* score quartile group for each CIN measurement, indicating that other mechanisms can also affect CIN. However, we observed a clear pattern that tumours with higher *HR* score indeed tend to have higher CIN level (Wilcoxon rank-sum test, one sided FDR p-value <0.05), with the exception of tumours in the third and fourth quartile groups in GII. Similar results were obtained for the remaining three cohorts (Supplementary Figures S8–S10). Overall, these results suggest an association between the extent of HR pathway dysregulation and the degree of CIN level in breast carcinomas.

As the *HR* score is based on gene expression, to ascertain whether the association observed above is due to the gene expression-based PAM50 subtypes, we performed the same analysis independently on tumours within each PAM50 subtype. In each analysis, the samples were divided into high

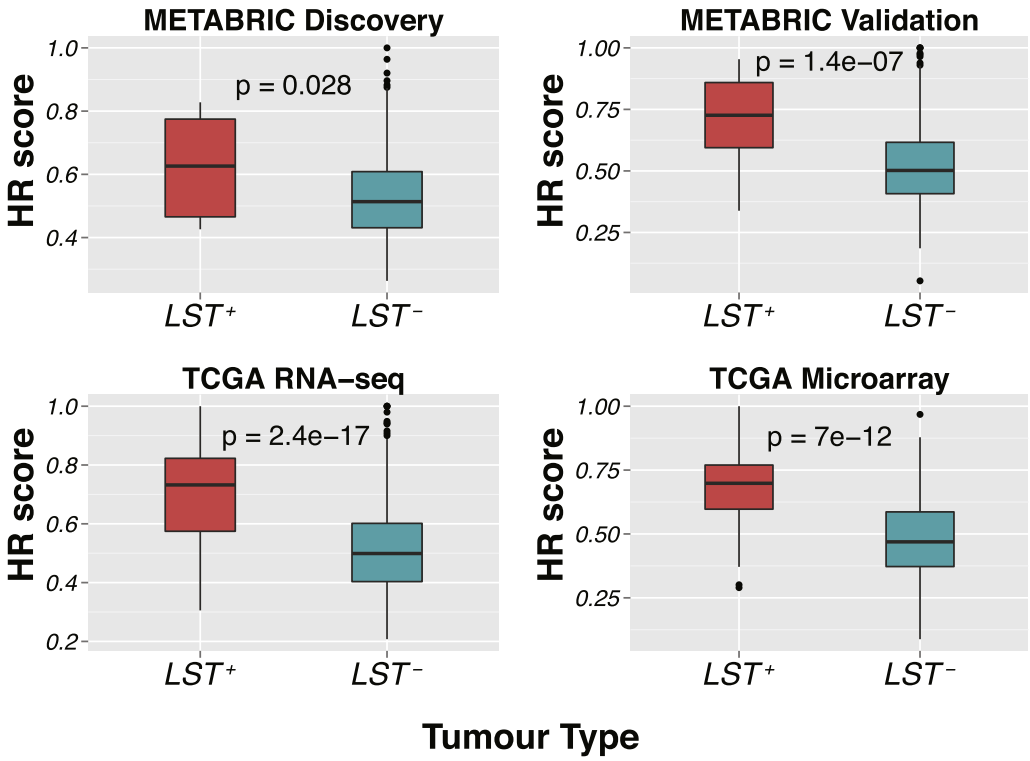

Figure 3 – Distribution of the *HR* score in LST$^+$ tumours and LST$^-$ tumours for each of the four cohorts. Colour represents LST status. The p-values were obtained using a Wilcoxon rank-sum test.

and low *HR* score groups according to the median. The results for the METABRIC Discovery cohort are summarised in Figure 6. For this cohort we consistently observed that tumours in the high *HR* score group have more breakpoints than do tumours in the low *HR* score group within the subtypes, despite the wide range of the breakpoint numbers observed for each subtype. The difference in GII between the low and high *HR* score groups was significant in Basal-like, Luminal A and Normal-like tumours, but not in HER2 and Luminal B tumours, while the difference in number of

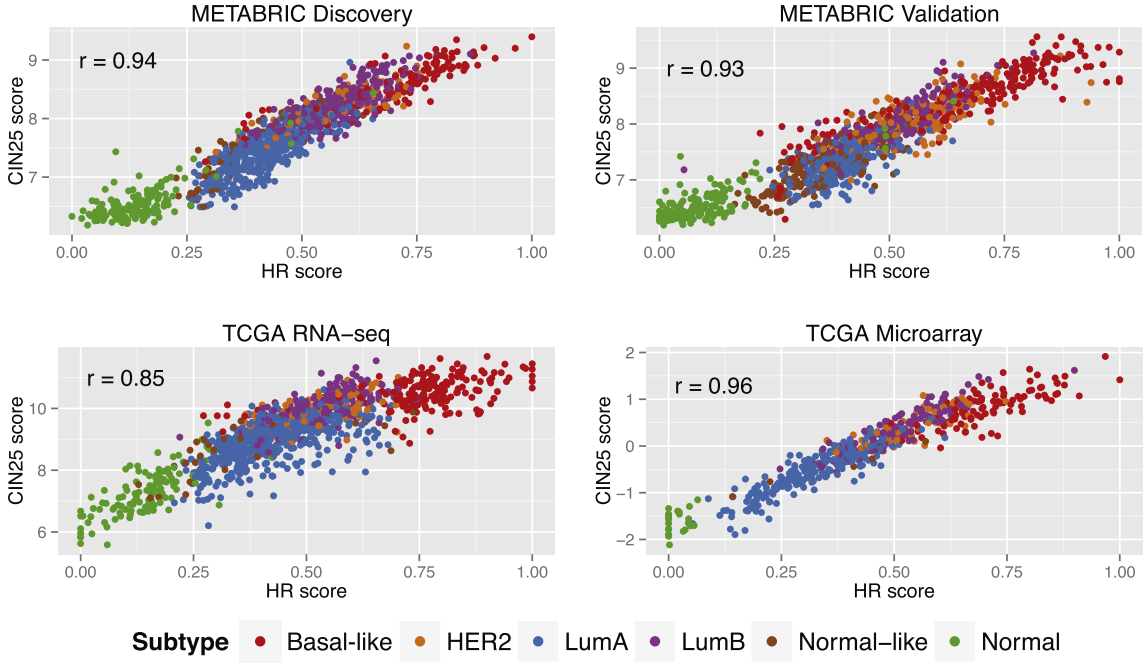

Figure 4 – Correlations between the CIN25 score and the *HR* score for each of the four cohorts.

amplified/deleted genes between the two groups was significant in all subtypes except HER2. For the other cohorts (Supplementary Figures S11–S13) we observed some differences between cohorts. For example, in the METABRIC Validation cohort, all three CIN measurements are significantly different between the two *HR* score groups for all subtypes, whereas the difference is significant in fewer subtypes in the TCGA Microarray cohort. These discrepancies might be due to low sample size in the TCGA Microarray cohort (e.g. there are only eight samples in its Normal-like subtype). Apart from these possible exceptions, the above results support the hypothesis that tumours with more-deregulated HR pathway are likely to have a higher degree of CIN, and this relationship can still be detected within the gene expression-based PAM50 subtypes.

3.3.3. *Association between the CIN measurements and other pathways*

The scatter plots in Figure 5 (METABRIC Discovery cohort) show that the *HR* score is moderately correlated with each of the three CIN measurements (breakpoints r = 0.60, GII r = 0.39 and number of amplified/deleted genes r = 0.48). These moderate correlations are not surprising, given that we do not consider aberrant HR as the only mechanism that contributes to CIN. In this section we investigated whether there are other pathways whose dysregulation also correlates with CIN, and whether these moderate correlations are far from random.

We computed a score for each of the 186 KEGG pathways (Kanehisa and Goto, 2000) and for 674 Reactome pathways (Croft et al., 2010), using the same approach as for the *HR* score. Spearman correlation coefficients between these scores and each of the three CIN measures were recorded and compared against the respective correlations between the *HR* score and the three CIN measurements. Figure 7 shows the results for the METABRIC Discovery cohort (KEGG pathways are in green and Reactome pathways in blue; similar results for the other three cohorts are in Supplementary Figures S14–S16). We found only a few KEGG or Reactome pathways whose dysregulation showed a similar level of correlation with CIN as did the HR pathway. For example, only four (2.2%) KEGG pathways (cell cycle, oocyte meiosis, progesterone-mediated oocyte maturation and p53 signalling) were more strongly associated with number of breakpoints than with the HR pathway (r = 0.61 − 0.63 compared to r = 0.60 for the HR pathway in Figure 7). Moreover, the strong associations of the oocyte meiosis, progesterone-mediated oocyte maturation and p53 signalling pathways with number of breakpoints is mainly due to their considerable overlap in gene content with the KEGG cell cycle pathway: 37%, 34% and 36% genes from each of these three pathways are also present in the cell cycle pathway (Supplementary Table S3). In contrast, only two HR genes are present in the cell cycle pathway. After removing the overlapping genes, association levels between each of these three pathways with number of breakpoints significantly decreased (results not shown). Similarly, although there were 24 (3.6%) Reactome pathways whose dysregulation showed a similar level of correlation with CIN as did the HR pathway, 18 of these are either the cell cycle pathway or its sub-pathways (Supplementary Table S4).

As the KEGG and Reactome pathways do not cover all genes measured in the whole-genome gene-expression profiling data analysed in this study, we also constructed 1000 "Random" pathways for each cohort to calculate an empirical p-value for the association between the HR score and each of the three CIN measurements. Each Random pathway is of the same length as HR but is composed of genes randomly selected from the gene-expression profiling data, excluding those from HR and cell cycle pathways. Similar to the KEGG pathways analysed above, we computed a score for each Random pathway, and compared the correlation coefficients with the three CIN measures against those for the *HR* score. As shown in Figure 7, only a few Random pathways (in pink) showed a level of association with CIN similar to that of the HR pathway, as indicated by the empirical p-values. Similar results for the other three cohorts were obtained (Supplementary Figures S14–S16).

Overall, these results indicate that the CIN level in tumours is associated with the dysregulation of only a limited number of pathways (e.g., the cell cycle pathway), and that the correlation between HR and CIN is far from being random.

3.4. *Association with survival in ER$^+$ tumours*

The two METABRIC cohorts are annotated with disease-specific survival data that are lacking for the two TCGA cohorts. We thus tested whether the *HR* score can predict patient survival in the two METABRIC cohorts. Figure 8 shows Kaplan–Maier plots for patients with ER$^+$ tumours from the METABRIC Discovery (n = 699; follow-up time ≤15 years) and validation cohorts (n = 582; follow-up time ≤15 years). For each cohort, patients were divided into high and low *HR* score groups based on the median *HR* score. For both cohorts, we observed a significant difference in patient survival between the two *HR* score groups with ER$^+$ tumours (Figure 8; Cox proportional hazards regression test p-value = 8.4e-04 and 3.9e-09 for the two cohorts, respectively). However, we observed no significant difference in survival between the two *HR* score groups for patients with ER$^-$ tumours (data not shown). As an association between CIN and prognosis in ER$^+$ tumours has already been documented (Przybytkowski et al., 2014; Smid et al., 2011), and after control for the number of chromosomal breaks there is no significant difference in survival between the two *HR* score-based groups (result not shown), we infer that the prognostic value of the *HR* score in ER$^+$ tumours is due to the association between the *HR* score and CIN.

4. Discussion

Multiple molecular mechanisms have been associated with the origin of CIN in cancer, including replication stress, telomere dysfunction, aberrant DNA repair and various defects in chromosome segregation (reviewed in Abbas et al., 2013; Aguilera and García-Muse, 2013; Negrini et al., 2010; Thompson et al., 2010). Although CIN can be experimentally induced by exploiting any of these mechanisms, replication stress has been recently identified as the first recurrent genetic defect associated with CIN in colorectal cancer (Burrell et al., 2013a). In this scenario, CIN is induced during DNA replication in fast-dividing tumour cells, giving rise to frequent

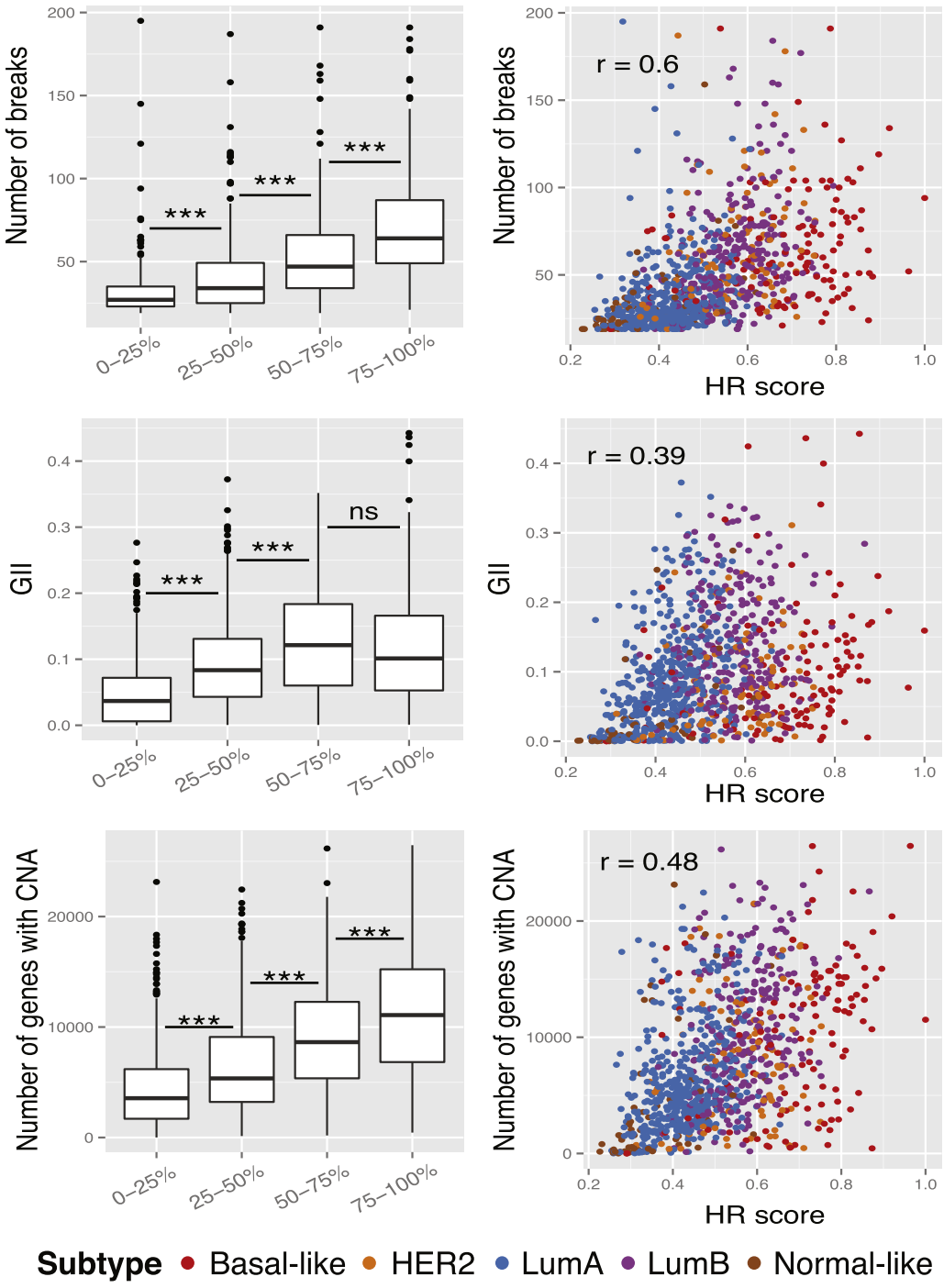

Figure 5 – *HR* score versus the three CIN measurements for the METABRIC Discovery cohort. Left: Boxplots of the three CIN measurements versus the four *HR* score quartile groups; stars indicate statistical significance according to a Wilcoxon rank-sum test: ns means not significant and * means p-value < 0.001. Right: Scatter plots of the *HR* score versus each of the three CIN measurements; r represents Pearson Correlation Coefficient.**

stalling of replication forks. Consequently, HR as the primary pathway for repair of the resultant DSBs during replication becomes overworked, and if HR is dysfunctional the frequency of replication stress-induced CIN is likely to increase dramatically. Here we have shown that HR dysregulation as measured by the *HR* score, which is indicative of aberrant HR repair, is prevalent in sporadic breast cancer and correlates with the level of CIN. We thus propose that HR dysregulation might contribute to replication stress-induced CIN at least in sporadic breast cancer. Consistent with this view, overexpression of the key HR gene *RAD51*, which is commonly seen in breast cancer as well as other cancer types, promotes chromosomal instability (Richardson et al., 2004), and two other critical HR genes, *BRCA1* and *BRCA2*, were recently proposed as

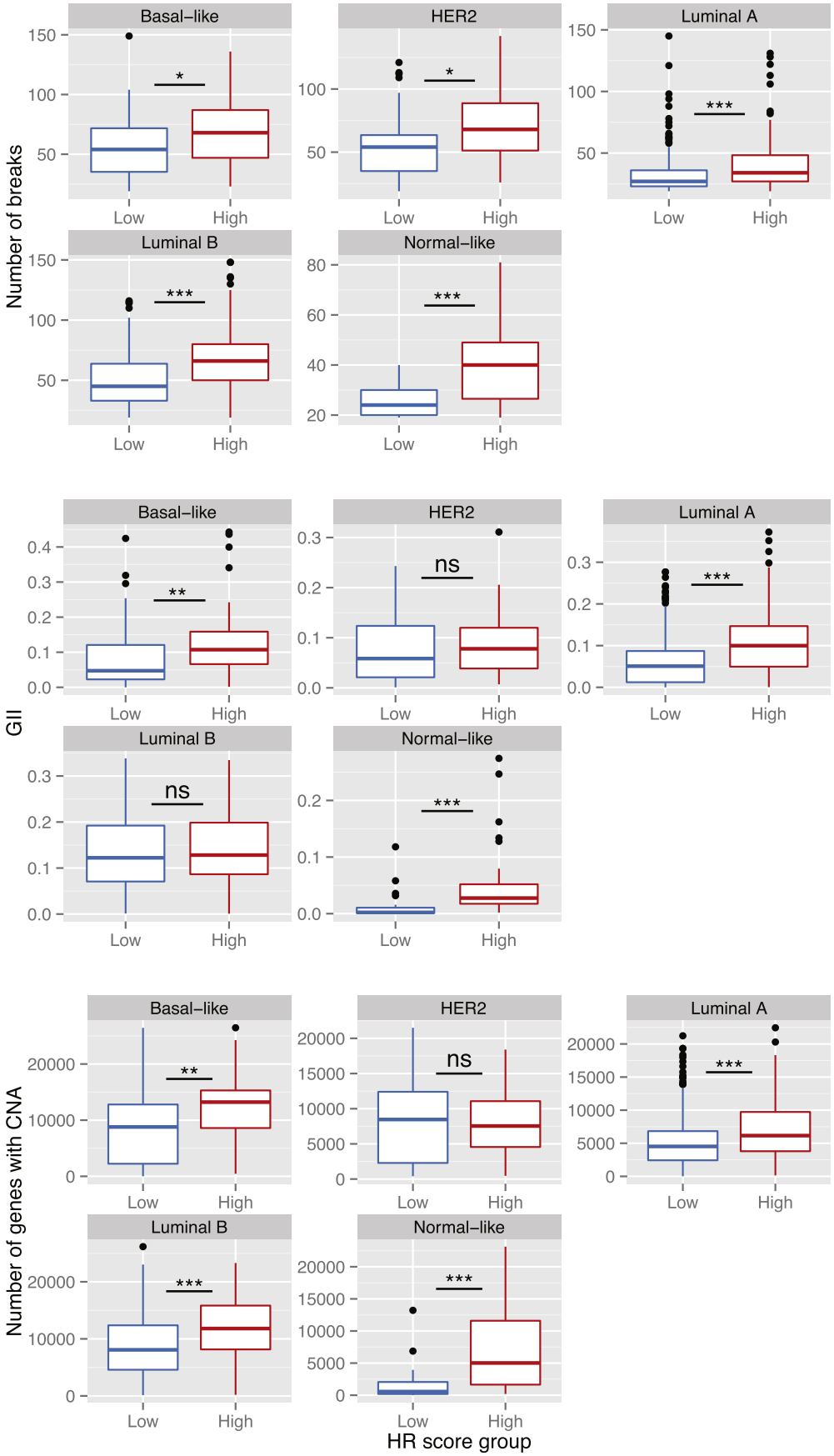

Figure 6 – *HR* score versus the three CIN measurements within PAM50 subtypes (METABRIC Discovery cohort). For each plot, the two *HR* score groups were divided according to the median *HR* score in each subtype; stars indicate the significance according to a Wilcoxon rank-sum test for each pair of groups: ns means not significant, * means $0.01 < p < 0.05$, ** means $0.001 < p < 0.01$, and * means $p < 0.001$.**

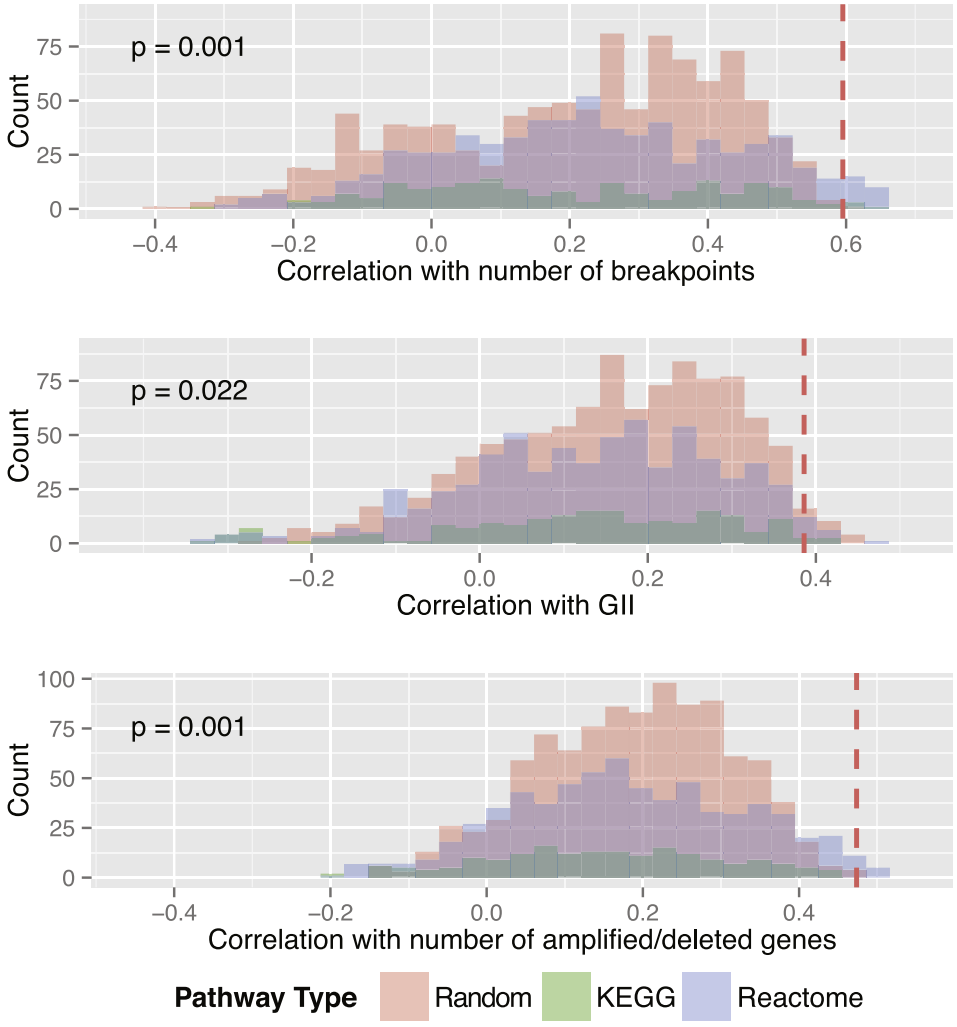

Figure 7 — Distributions of the correlations between pathway scores and the three CIN measurements (METABRIC Discovery cohort). Results for KEGG pathways are in green, Reactome pathways in blue and Random pathways in pink. Spearman correlation coefficients (r) are represented on the x-axis. Pathway score were calculated with Pathifier. The vertical dashed line in each histogram indicates the value of r between the *HR* score and each of the three CIN measurements, and p represents an empirical p-value for that value of r.

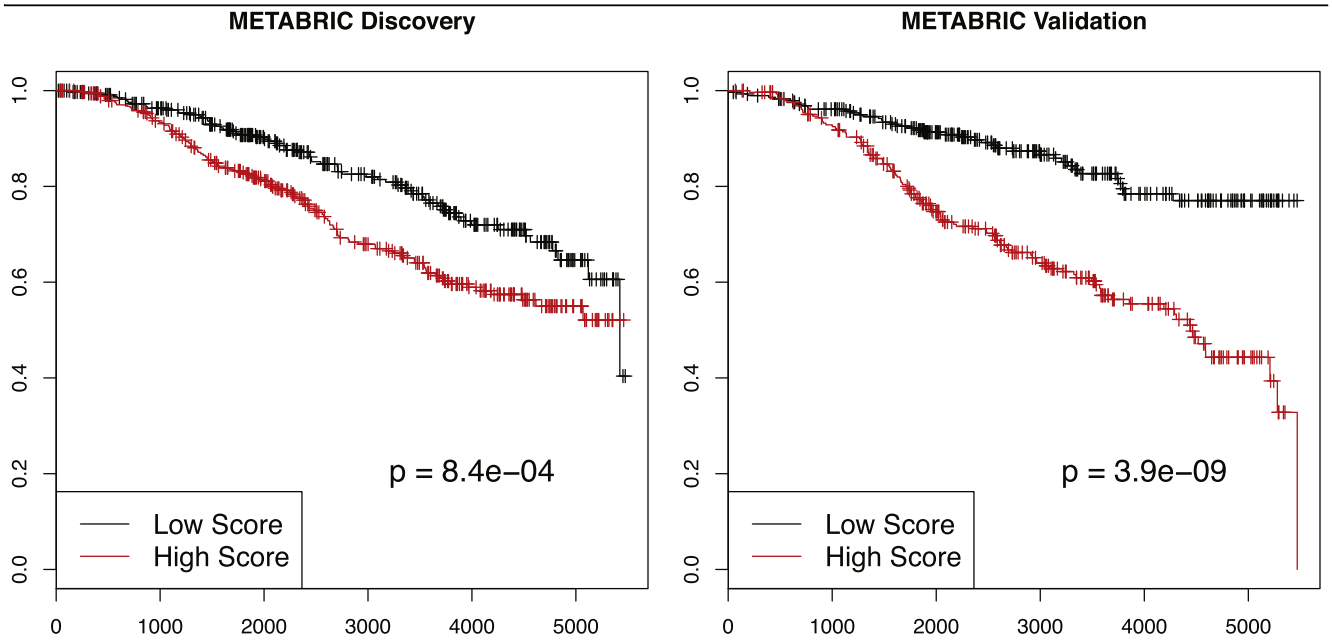

Figure 8 — Kaplan–Maier plot for disease specific survival in the METABRIC Discovery cohort (left) and Validation cohort (right). Patients with ER$^+$ tumour were divided into two equal-sized groups based on the median *HR* score in each cohort.

chromosome custodians mainly due to their role in HR (Venkitaraman, 2014a, b).

Dysfunction of the HR pathway, although not the primary cause, may increase the level of replication stress-induced CIN in several ways. Firstly, it can cause inefficient repair of DSBs, resulting in an accumulation of chromosomal breaks. Secondly, by triggering error-prone repair pathways including canonical non-homologous end-joining (C-NHEJ) and alternative non-homologous end-joining (Alt-NHEJ, also called microhomology-mediated end joining (MMEJ)), HR dysfunction can lead to translocations, translocation-related chromosomal breaks and DNA copy-number changes. Specifically, in contrast to HR that requires homologous sequence to guide repair, C-NHEJ and Alt-NHEJ mediate the repair by a direct ligation of the break ends after more-or-less end processing, and so do not ensure that the broken DNA strands are rejoined in the correct position. These two low-fidelity pathways come to repair DSBs generated during DNA replication when HR is deficient, resulting in translocation as well as translocation-related chromosomal breaks (Alexandrov et al., 2013; Bunting and Nussenzweig, 2013; Ottaviani et al., 2014; Villarreal et al., 2012). Moreover, gene copy number changes also arise when the repair of broken replication forks switched from HR to the two NHEJs, especially Alt-NHEJ (Hastings et al., 2009);

A third way in which HR pathway dysfunction can increase replication stress-induced CIN is by affecting mitosis and the proper functioning of telomeres. HR defects and the consequent slow progression of replication forks can elicit alterations of mitosis, which highlights the importance of HR at the interface of these two processes for protection against CIN (Wilhelm et al., 2014). In addition, DSB repair is shut down during the M phase to avoid telomere fusion and as a consequence, mitosis will continue even in the presence of DSBs or fragmented chromosomes, giving rise to CIN (Orthwein et al., 2014). This emphasises the importance of DSB repair during DNA replication, especially given the presence of DSBs that result from replication stress. HR defects caused by *BRCA2* mutations could also lead to telomere dysfunction, a mechanism that has been proposed to explain, in part, the chromosomal instability observed in *BRCA2*-deficient tumours (Badie et al., 2010). Taken together, HR dysfunction can increase CIN via diverse mechanisms, and the association revealed in this study between HR dysregulation and CIN (Figures 4–6) indicates that dysregulated HR might contribute to the CIN observed in highly replicative tumours.

The study of CIN in breast cancer has attracted immense interest in recent years following the recognition of its clinical relevance in disease heterogeneity, drug resistance and patient response (A'Hern et al., 2013; Birkbak et al., 2011; Endesfelder et al., 2014; Habermann et al., 2009; Roylance et al., 2011; Sansregret and Nepveu, 2011; Swanton et al., 2009; Vincent-Salomon et al., 2013); reviewed by (Wiechec, 2011). CIN induces evolution in tumours, providing the heterogeneity from which aggressive and/or drug-resistant tumour clones are selectively established. CIN aids tumour development by amplifying genomic regions containing oncogenes and deleting regions containing tumour-suppressor genes, thereby significantly influencing treatment response and survival in patients. Our results further strengthen this connection by associating dysregulated HR with the extent of amplified/deleted genes and regions of the chromosome, and by showing that ER^+ tumours with high HR score or CIN levels display significantly poorer prognosis (Figure 8).

A measure of HR dysregulation such as the one adopted here can be extremely valuable to guide therapeutic options. The observation that cancer cells deficient in HR are profoundly sensitive to PARP inhibitors (Bryant et al., 2005; Farmer et al., 2005) has already led to the development of targeted PARP therapies for sporadic breast and ovarian cancers with defects in core HR genes such as *BRCA1* and *BRCA2*, a condition termed as "BRCAness" (Turner et al., 2004). PARP is an important protein family whose members function in restarting stalled replication forks and diverting DSBs to HR-mediated repair. It has been proposed that accumulated chromosomal instability arising from the continued stalling of replication forks, accompanied by deficiency in repairing DSBs and thereby triggering a genomic catastrophe, may explain how PARP inhibition kills HR-deficient cancer cells (Bryant et al., 2005; Farmer et al., 2005). Although focussing on a mechanistic explanation for PARP-based cancer therapy, these models indirectly suggest an underlying relationship among replicative stress, dysfunctional HR and the accumulation of chromosomal instability.

In conclusion, we performed a personalised pathway analysis by calculating an *HR* score that quantifies HR pathway dysregulation in individual breast tumours, with the behaviour of HR in normal breast tissues serving as a benchmark. Our results are reproducible across four large breast cancer cohorts (~3000 tumours in total). We found HR is dysregulated to various extents between and within the gene expression-based PAM50 subtypes, which may reflect their HR repair deficiency. More importantly, we uncovered a novel association between HR dysregulation and CIN. Although HR has a well-known role in maintaining genomic integrity, this work is the first large-scale study to assess the correlation between HR dysregulation and CIN in sporadic breast cancer. As such our results will be useful for future studies that aim to identify causative factors of CIN in sporadic breast cancer as well as in other cancer types.

Acknowledgements

This study makes use of data generated by the Molecular Taxonomy of Breast Cancer International Consortium funded by Cancer Research UK and the British Columbia Cancer Agency Branch. We also thank TCGA for providing the genomic data. This study was funded by the Australian National Health and Medical Research Council (NHMRC) Project Grant (ID: 1028742) to PTS and MAR. KALC was supported in part by the Australian Cancer Research Foundation (ACRF) for the Diamantina Individualised Oncology Care Centre at The University of Queensland Diamantina Institute and the NHMRC Career Development fellowship (ID: 1087415). KKK is an NHMRC Senior Principal Search Fellow (ID: 613638) supported by the NHMRC Project Grant (ID: 1017028).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2015.09.007.

REFERENCES

A'Hern, R.P., Jamal-Hanjani, M., Szász, A.M., Johnston, S.R.D., Reis-Filho, J.S., Roylance, R., Swanton, C., 2013. Taxane benefit in breast cancer–a role for grade and chromosomal stability. Nat. Rev. Clin. Oncol. 10, 357–364. http://dx.doi.org/10.1038/nrclinonc.2013.67.

Abbas, T., Keaton, M.A., Dutta, A., 2013. Genomic instability in cancer. Cold Spring Harbor Perspect. Biol. 5, a012914. http://dx.doi.org/10.1101/cshperspect.a012914.

Aguilera, A., García-Muse, T., 2013. Causes of genome instability. Annu. Rev. Genet. 47, 1–32. http://dx.doi.org/10.1146/annurev-genet-111212-133232.

Ahn, T., Lee, E., Huh, N., Park, T., 2014. Personalized identification of altered pathways in cancer using accumulated normal tissue data. Bioinformatics 30, i422–i429. http://dx.doi.org/10.1093/bioinformatics/btu449.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al., 2013. Signatures of mutational processes in human cancer. Nature 500, 415–421. http://dx.doi.org/10.1038/nature12477.

Badie, S., Escandell, J.M., Bouwman, P., Carlos, A.R., Thanasoula, M., Gallardo, M.M., Suram, A., Jaco, I., Benitez, J., Herbig, U., Blasco, M.A., Jonkers, J., Tarsounas, M., 2010. BRCA2 acts as a RAD51 loader to facilitate telomere replication and capping. Nat. Struct. Mol. Biol. 17, 1461–1469. http://dx.doi.org/10.1038/nsmb.1943.

Bekker-Jensen, S., Mailand, N., 2011. The ubiquitin- and SUMO-dependent signaling response to DNA double-strand breaks. FEBS Lett. 585, 2914–2919. http://dx.doi.org/10.1016/j.febslet.2011.05.056.

Birkbak, N.J., Eklund, A.C., Li, Q., McClelland, S.E., Endesfelder, D., Tan, P., Tan, I.B., Richardson, A.L., Szallasi, Z., Swanton, C., 2011. Paradoxical relationship between chromosomal instability and survival outcome in cancer. Cancer Res. 71, 3447–3452. http://dx.doi.org/10.1158/0008-5472.CAN-10-3667.

Brown, E.T., Holt, J.T., 2009. Rad51 overexpression rescues radiation resistance in BRCA2-defective cancer cells. Mol. Carcinog. 48, 105–109. http://dx.doi.org/10.1002/mc.20463.

Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., Helleday, T., 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature 434, 913–917. http://dx.doi.org/10.1038/nature03443.

Bunting, S.F., Nussenzweig, A., 2013. End-joining, translocations and cancer. Nat. Rev. Cancer 13, 443–454. http://dx.doi.org/10.1038/nrc3537.

Burrell, R.A., McClelland, S.E., Endesfelder, D., Groth, P., Weller, M.-C., Shaikh, N., Domingo, E., Kanu, N., Dewhurst, S.M., Gronroos, E., Chew, S.K., Rowan, A.J., Schenk, A., Sheffer, M., Howell, M., Kschischo, M., Behrens, A., Helleday, T., Bartek, J., Tomlinson, I.P., Swanton, C., 2013a. Replication stress links structural and numerical cancer chromosomal instability. Nature 494, 492–496. http://dx.doi.org/10.1038/nature11935.

Burrell, R.A., McGranahan, N., Bartek, J., Swanton, C., 2013b. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 501, 338–345. http://dx.doi.org/10.1038/nature12625.

Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., Szallasi, Z., 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat. Genet. 38, 1043–1048. http://dx.doi.org/10.1038/ng1861.

Chowdhury, D., Choi, Y.-E., Brault, M.E., 2013. Charity begins at home: non-coding RNA functions in DNA repair. Nat. Rev. Mol. Cell Biol. 14, 181–189. http://dx.doi.org/10.1038/nrm3523.

Ciriello, G., Sinha, R., Hoadley, K.A., Jacobsen, A.S., Reva, B., Perou, C.M., Sander, C., Schultz, N., 2013. The molecular diversity of Luminal A breast tumors. Breast Cancer Res. Treat 141, 409–420. http://dx.doi.org/10.1007/s10549-013-2699-3.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., Stein, L., 2010. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 39, D691–D697. http://dx.doi.org/10.1093/nar/gkq1018.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., et al., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, 1–7. http://dx.doi.org/10.1038/nature10983.

d'Adda di Fagagna, F., 2014. A direct role for small non-coding RNAs in DNA damage response. Trends Cell Biol. 24, 171–178. http://dx.doi.org/10.1016/j.tcb.2013.09.008.

Davis, J.D., 2011. DNA damage and breast cancer. WJCO 2, 329. http://dx.doi.org/10.5306/wjco.v2.i9.329.

Dou, H., Huang, C., Van Nguyen, T., Lu, L.-S., Yeh, E.T.H., 2011. SUMOylation and de-SUMOylation in response to DNA damage. FEBS Lett. 585, 2891–2896. http://dx.doi.org/10.1016/j.febslet.2011.04.002.

Drier, Y., Sheffer, M., Domany, E., 2013. Pathway-based personalized analysis of cancer. Proc. Natl. Acad. Sci. U. S. A 110, 6388–6393. http://dx.doi.org/10.1073/pnas.1219651110.

Endesfelder, D., Burrell, R.A., Kanu, N., McGranahan, N., Howell, M., Parker, P.J., Downward, J., Swanton, C., Kschischo, M., 2014. Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in ER^+ breast cancer. Cancer Res. 74, 4853–4863. http://dx.doi.org/10.1158/0008-5472.CAN-13-2664.

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., Martin, N.M.B., Jackson, S.P., Smith, G.C.M., Ashworth, A., 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434, 917–921. http://dx.doi.org/10.1038/nature03445.

Habermann, J.K., Doering, J., Hautaniemi, S., Roblick, U.J., Bündgen, N.K., Nicorici, D., Kronenwett, U., Rathnagiriswaran, S., Mettu, R.K.R., Ma, Y., Krüger, S., Bruch, H.-P., Auer, G., Guo, N.L., Ried, T., 2009. The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. Int. J. Cancer 124, 1552–1564. http://dx.doi.org/10.1002/ijc.24017.

Hastings, P.J., Lupski, J.R., Rosenberg, S.M., Ira, G., 2009. Mechanisms of change in gene copy number. Nat. Rev. Genet. 10, 551–564. http://dx.doi.org/10.1038/nrg2593.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M.G., Sawyer, L.R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J.P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J.F., Ellis, M.J., Olopade, O.I., Bernard, P.S., Perou, C.M., 2006. The molecular portraits of breast tumors are

conserved across microarray platforms. BMC Genomics 7, 96. http://dx.doi.org/10.1186/1471-2164-7-96.

Huang, S., Yee, C., Ching, T., Yu, H., Garmire, L.X., 2014. A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer. PLoS Comput. Biol. 10, e1003851. http://dx.doi.org/10.1371/journal.pcbi.1003851.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., et al., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70. http://dx.doi.org/10.1038/nature11412.

Lee, A.J.X., Endesfelder, D., Rowan, A.J., Walther, A., Birkbak, N.J., Futreal, P.A., Downward, J., Szallasi, Z., Tomlinson, I.P.M., Howell, M., Kschischo, M., Swanton, C., 2011. Chromosomal instability confers intrinsic multidrug resistance. Cancer Res. 71, 1858–1870. http://dx.doi.org/10.1158/0008-5472.CAN-10-3604.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., Pietenpol, J.A., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121, 2750–2767. http://dx.doi.org/10.1172/JCI45014.

Liu, C., Srihari, S., Lê Cao, K.-A., Chenevix-Trench, G., Simpson, P.T., Ragan, M.A., Khanna, K.K., 2014. A fine-scale dissection of the DNA double-strand break repair machinery and its implications for breast cancer therapy. Nucleic Acids Res. 42, 6106–6127. http://dx.doi.org/10.1093/nar/gku284.

Livshits, A., Git, A., Fuks, G., Caldas, C., Domany, E., 2015. Pathway-based personalized analysis of breast cancer expression data. Mol. Oncol 9, 1471–1483. http://dx.doi.org/10.1016/j.molonc.2015.04.006.

Martin, R.W., Orelli, B.J., Yamazoe, M., Minn, A.J., Takeda, S., Bishop, D.K., 2007. RAD51 up-regulation bypasses BRCA1 function and is a common feature of BRCA1-deficient breast tumors. Cancer Res. 67, 9658–9665. http://dx.doi.org/10.1158/0008-5472.CAN-07-0290.

Negrini, S., Gorgoulis, V.G., Halazonetis, T.D., 2010. Genomic instability—an evolving hallmark of cancer. Nat. Rev. Mol. Cell Biol 11, 220–228. http://dx.doi.org/10.1038/nrm2858.

Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M., Strecker, J., Escribano-Diaz, C., Durocher, D., 2014. Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. Science 344, 189–193. http://dx.doi.org/10.1126/science.1248024.

Ottaviani, D., LeCain, M., Sheer, D., 2014. The role of microhomology in genomic structural variation. Trends Genet. 30, 85–94. http://dx.doi.org/10.1016/j.tig.2014.01.001.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J.F., Stijleman, I.J., Palazzo, J., Marron, J.S., Nobel, A.B., Mardis, E., Nielsen, T.O., Ellis, M.J., Perou, C.M., Bernard, P.S., 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. 27, 1160–1167. http://dx.doi.org/10.1200/JCO.2008.18.1370.

Peng, G., Chun-Jen Lin, C., Mo, W., Dai, H., Park, Y.-Y., Kim, S.M., Peng, Y., Mo, Q., Siwko, S., Hu, R., Lee, J.-S., Hennessy, B., Hanash, S., Mills, G.B., Lin, S.-Y., 2014. Genome-wide transcriptome profiling of homologous recombination DNA repair. Nat. Commun. 5, 3361. http://dx.doi.org/10.1038/ncomms4361.

Peng, Y., Dai, H., Wang, E., Lin, C.C.-J., Mo, W., Peng, G., Lin, S.-Y., 2015. TUSC4 functions as a tumor suppressor by regulating BRCA1 stability. Cancer Res. 75, 378–386. http://dx.doi.org/10.1158/0008-5472.CAN-14-2315.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lønning, P.E., Borresen-Dale, A.L., Brown, P.O., Botstein, D., 2000. Molecular portraits of human breast tumours. Nature 406, 747–752. http://dx.doi.org/10.1038/35021093.

Pitroda, S.P., Pashtan, I.M., Logan, H.L., Budke, B., Darga, T.E., Weichselbaum, R.R., Connell, P.P., 2014. DNA repair pathway gene expression score correlates with repair proficiency and tumor sensitivity to chemotherapy. Sci. Transl Med. 6, 229ra42. http://dx.doi.org/10.1126/scitranslmed.3008291.

Popova, T., Manié, E., Rieunier, G., Caux-Moncoutier, V., Tirapo, C., Dubois, T., Delattre, O., Sigal-Zafrani, B., Bollet, M., Longy, M., Houdayer, C., Sastre-Garau, X., Vincent-Salomon, A., Stoppa-Lyonnet, D., Stern, M.-H., 2012. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. Cancer Res. 72, 5454–5462. http://dx.doi.org/10.1158/0008-5472.CAN-12-1470.

Price, B.D., D'Andrea, A.D., 2013. Chromatin remodeling at DNA double-strand breaks. Cell 152, 1344–1354. http://dx.doi.org/10.1016/j.cell.2013.02.011.

Przybytkowski, E., Lenkiewicz, E., Barrett, M.T., Klein, K., Nabavi, S., Greenwood, C.M.T., Basik, M., 2014. Chromosome-breakage genomic instability and chromothripsis in breast cancer. BMC Genomics 15, 579. http://dx.doi.org/10.1186/1471-2164-15-579.

Richardson, C., Stark, J.M., Ommundsen, M., Jasin, M., 2004. Rad51 overexpression promotes alternative double-strand break repair pathways and genome instability. Oncogene 23, 546–553. http://dx.doi.org/10.1038/sj.onc.1207098.

Roylance, R., Endesfelder, D., Gorman, P., Burrell, R.A., Sander, J., Tomlinson, I., Hanby, A.M., Speirs, V., Richardson, A.L., Birkbak, N.J., Eklund, A.C., Downward, J., Kschischo, M., Szallasi, Z., Swanton, C., 2011. Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. Cancer Epidemiol. Biomarkers Prev. 20, 2183–2194. http://dx.doi.org/10.1158/1055-9965.EPI-11-0343.

Sansregret, L., Nepveu, A., 2011. Gene signatures of genomic instability as prognostic tools for breast cancer. Future Oncol. 7, 591–594. http://dx.doi.org/10.2217/fon.11.34.

Sarasin, A., Kauffmann, A., 2008. Overexpression of DNA repair genes is associated with metastasis: a new hypothesis. Mutat. Res. 659, 49–55. http://dx.doi.org/10.1016/j.mrrev.2007.12.002.

Sharma, V., Misteli, T., 2013. Non-coding RNAs in DNA damage and repair. FEBS Lett. 587, 1832–1839. http://dx.doi.org/10.1016/j.febslet.2013.05.006.

Smid, M., Hoes, M., Sieuwerts, A.M., Sleijfer, S., Zhang, Y., Wang, Y., Foekens, J.A., Martens, J.W.M., 2011. Patterns and incidence of chromosomal instability and their prognostic relevance in breast cancer subtypes. Breast Cancer Res. Treat 128, 23–30. http://dx.doi.org/10.1007/s10549-010-1026-5.

Swanton, C., Nicke, B., Schuett, M., Eklund, A.C., Ng, C., Li, Q., Hardcastle, T., Lee, A., Roy, R., East, P., Kschischo, M., Endesfelder, D., Wylie, P., Kim, S.N., Chen, J.-G., Howell, M., Ried, T., Habermann, J.K., Auer, G., Brenton, J.D., Szallasi, Z., Downward, J., 2009. Chromosomal instability determines taxane response. Proc. Natl. Acad. Sci. U. S. A 106, 8671–8676. http://dx.doi.org/10.1073/pnas.0811835106.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Lønning, P.E., Borresen-Dale, A.L., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A 98, 10869–10874. http://dx.doi.org/10.1073/pnas.191367098.

Takata, M., Sasaki, M.S., Tachiiri, S., Fukushima, T., Sonoda, E., Schild, D., Thompson, L.H., Takeda, S., 2001. Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. Mol. Cell Biol. 21, 2858–2866. http://dx.doi.org/10.1128/MCB.21.8.2858-2866.2001.

Thompson, S.L., Bakhoum, S.F., Compton, D.A., 2010. Mechanisms of chromosomal instability. Curr. Biol. 20, R285–R295. http://dx.doi.org/10.1016/j.cub.2010.01.034.

Turner, N., Tutt, A., Ashworth, A., 2004. Hallmarks of "BRCAness" in sporadic cancers. Nat. Rev. Cancer 4, 814–819. http://dx.doi.org/10.1038/nrc1457.

Ulrich, H.D., 2012. Ubiquitin and SUMO in DNA repair at a glance. J. Cell Sci. 125, 249–254. http://dx.doi.org/10.1242/jcs.091801.

van Attikum, H., Gasser, S.M., 2009. Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol. 19, 207–217. http://dx.doi.org/10.1016/j.tcb.2009.03.001.

Vaske, C.J., Benz, S.C., Sanborn, J.Z., Earl, D., Szeto, C., Zhu, J., Haussler, D., Stuart, J.M., 2010. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics 26, i237–i245. http://dx.doi.org/10.1093/bioinformatics/btq182.

Venkitaraman, A.R., 2014a. Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. Science 343, 1470–1475. http://dx.doi.org/10.1126/science.1252230.

Venkitaraman, A.R., 2014b. Tumour suppressor mechanisms in the control of chromosome stability: insights from BRCA2. Mol. Cells 37, 95–99. http://dx.doi.org/10.14348/molcells.2014.2346.

Villarreal, D.D., Lee, K., Deem, A., Shim, E.Y., Malkova, A., Lee, S.E., 2012. Microhomology directs diverse DNA break repair pathways and chromosomal translocations. PLoS Genet. 8, e1003026. http://dx.doi.org/10.1371/journal.pgen.1003026.

Vincent-Salomon, A., Benhamo, V., Gravier, E., Rigaill, G., Gruel, N., Robin, S., de Rycke, Y., Mariani, O., Pierron, G., Gentien, D., Reyal, F., Cottu, P., Fourquet, A., Rouzier, R., Sastre-Garau, X., Delattre, O., 2013. Genomic instability: a stronger prognostic marker than proliferation for early stage luminal breast carcinomas. PLoS One 8, e76496. http://dx.doi.org/10.1371/journal.pone.0076496.

Vollan, H.K.M., Rueda, O.M., Chin, S.-F., Curtis, C., Turashvili, G., Shah, S., Lingjaerde, O.C., Yuan, Y., Ng, C.K., Dunning, M.J., Dicks, E., Provenzano, E., Sammut, S., McKinney, S., Ellis, I.O., Pinder, S., Purushotham, A., Murphy, L.C., Kristensen, V.N., METABRIC Group, Brenton, J.D., Pharoah, P.D.P., Børresen-Dale, A.-L., Aparicio, S., Caldas, C., 2015. A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. Mol. Oncol. 9, 115–127. http://dx.doi.org/10.1016/j.molonc.2014.07.019.

Wang, H., Cai, H., Ao, L., Yan, H., Zhao, W., Qi, L., Gu, Y., Guo, Z., 2015a. Individualized identification of disease-associated pathways with disrupted coordination of gene expression. Brief. Bioinform.. http://dx.doi.org/10.1093/bib/bbv030

Wang, H., Sun, Q., Zhao, W., Qi, L., Gu, Y., Li, P., Zhang, M., Li, Y., Liu, S.-L., Guo, Z., 2015b. Individual-level analysis of differential expression of genes and pathways for personalized medicine. Bioinformatics 31, 62–68. http://dx.doi.org/10.1093/bioinformatics/btu522.

Watkins, J., Weekes, D., Shah, V., Gazinska, P., Joshi, S., Sidhu, B., Gillett, C., Pinder, S., Vanoli, F., Jasin, M., Mayrhofer, M., Isaksson, A., Cheang, M.C.U., Mirza, H., Frankum, J., Lord, C.J., Ashworth, A., Vinayak, S., Ford, J.M., Telli, M.L., Grigoriadis, A., Tutt, A.N.J., 2015. Genomic complexity profiling reveals that HORMAD1 overexpression contributes to homologous recombination deficiency in triple-negative breast cancers. Cancer Discov. 5, 488–505. http://dx.doi.org/10.1158/2159-8290.CD-14-1092.

Wiechec, E., 2011. Implications of genomic instability in the diagnosis and treatment of breast cancer. Expert Rev. Mol. Diagn. 11, 445–453. http://dx.doi.org/10.1586/erm.11.21.

Wilhelm, T., Magdalou, I., Barascu, A., Técher, H., Debatisse, M., Lopez, B.S., 2014. Spontaneous slow replication fork progression elicits mitosis alterations in homologous recombination-deficient mammalian cells. Proc. Natl. Acad. Sci. U. S. A 111, 763–768. http://dx.doi.org/10.1073/pnas.1311520111.

Yanagawa, M., Ikemot, K., Kawauchi, S., Furuya, T., Yamamoto, S., Oka, M., Oga, A., Nagashima, Y., Sasaki, K., 2012. Luminal A and luminal B (HER2 negative) subtypes of breast cancer consist of a mixture of tumors with different genotype. BMC Res. Notes 5, 376. http://dx.doi.org/10.1186/1756-0500-5-376.

Zeman, M.K., Cimprich, K.A., 2014. Causes and consequences of replication stress. Nat. Cell Biol. 16, 2–9. http://dx.doi.org/10.1038/ncb2897.

Zheng, L., Dai, H., Zhou, M., Li, X., Liu, C., Guo, Z., Wu, X., Wu, J., Wang, C., Zhong, J., Huang, Q., Garcia-Aguilar, J., Pfeifer, G.P., Shen, B., 2012. Polyploid cells rewire DNA damage response networks to overcome replication stress-induced barriers for tumour progression. Nat. Commun. 3, 815. http://dx.doi.org/10.1038/ncomms1825.