An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers

Rachael Natrajan, Britta Weigelt, Alan Mackay, Felipe C. Geyer, Anita Grigoriadis, David S. P. Tan, Chris Jones, Christopher J. Lord, Radost Vatcheva, Socorro M. Rodriguez-Pinilla, et al.

Breast Cancer Res Treat (2010) 121:575–589
DOI 10.1007/s10549-009-0501-3

PRECLINICAL STUDY

An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers

Rachael Natrajan · Britta Weigelt · Alan Mackay · Felipe C. Geyer · Anita Grigoriadis · David S. P. Tan · Chris Jones · Christopher J. Lord · Radost Vatcheva · Socorro M. Rodriguez-Pinilla · Jose Palacios · Alan Ashworth · Jorge S. Reis-Filho

Received: 24 July 2009 / Accepted: 28 July 2009 / Published online: 18 August 2009

Abstract Breast cancer is a heterogeneous disease caused by the accumulation of genetic changes in neoplastic cells. We hypothesised that molecular subtypes of breast cancer may be driven by specific constellations of genes whose expression is regulated by gene copy number aberrations. To address this question, we analysed a series of 48 microdissected grade III ductal carcinomas using high resolution microarray comparative genomic hybridisation and mRNA expression arrays. There were 5,931 genes whose expression significantly correlates with copy number identified; out of these, 1,897 genes were significantly differentially expressed between basal-like, HER2 and luminal tumours. Ingenuity Pathway Analysis (IPA) revealed that 'G1/S cell cycle regulation' and 'BRCA1 in DNA damage control' pathways were significantly enriched for genes whose expression correlates with copy number and are differentially expressed between the molecular subtypes of breast cancer. IPA of genes whose expression significantly correlates with copy number in each molecular subtype individually revealed that canonical pathways involved in oestrogen receptor (ER) signalling and DNA repair are enriched for these genes. We also identified 32, 157 and 265 genes significantly overexpressed when amplified in basal-like, HER2 and luminal cancers, respectively. These lists include known and novel potential therapeutic targets (e.g. *HER2* and *PPM1D* in HER2 cancers). Our results provide strong circumstantial evidence that different patterns of genetic aberrations in distinct molecular subtypes of breast cancer contribute to their specific transcriptomic profiles and that biological phenomena characteristic of each subtype (e.g. proliferation, HER2 and ER signalling) may be driven by specific patterns of copy number aberrations.

Keywords Breast cancer · Molecular subtypes · Copy number · Gene expression · Data integration · Molecular pathways

Rachael Natrajan and Britta Weigelt contributed equally to the present study.

Electronic supplementary material The online version of this article (doi:10.1007/s10549-009-0501-3) contains supplementary material, which is available to authorized users.

R. Natrajan (✉) · A. Mackay · F. C. Geyer · D. S. P. Tan · C. J. Lord · R. Vatcheva · A. Ashworth · J. S. Reis-Filho (✉)
The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK
e-mail: Rachael.Natrajan@icr.ac.uk

J. S. Reis-Filho
e-mail: Jorge.Reis-Filho@icr.ac.uk

B. Weigelt
Cancer Research UK, London Research Institute, London WC2A 3PX, UK

A. Grigoriadis
Breakthrough Breast Research Unit, Guy's Hospital, London SE1 9RT, UK

C. Jones
Paediatric Oncology, Institute of Cancer Research, Sutton SM2 5NG, UK

S. M. Rodriguez-Pinilla
Centro Nacional de Investigaciones Oncologicas (CNIO), Madrid, Spain

J. Palacios
Servicio de Anatomia Patologica, Hospital Virgen del Rocío, Seville, Spain

Introduction

Breast cancer is a heterogeneous disease comprising multiple entities, which have distinct biological features and clinical behaviour. In recent years, it has become apparent that breast cancer heterogeneity may be underpinned by distinct genomic changes [1–7]. Microarray-based gene expression profiling of breast cancer has led to the identification of at least five distinct classes of breast cancers: luminal A, luminal B, HER2, normal breast-like and basal-like breast cancer. It has been hypothesised that the differences between these molecular subgroups of breast cancer stem from differences in their cell of origin [8, 9]. However, luminal, HER2 and basal-like breast cancers have been shown to harbour discrete patterns of genetic aberrations, which may also contribute to their distinct transcriptomic profiles and clinical phenotypes [1–3, 5, 10].

Gene copy number aberrations have been observed in >95% of breast cancers and involve a substantial proportion of the genome of breast cancer cells [1–3, 5, 6, 10–12]. It is unclear, however, as to what extent copy number aberrations drive the expression of genes and molecular networks and pathways in breast cancer. Previous efforts in identifying these networks and pathways were either carried out in breast cancer cell lines [13] or made use of non-microdissected tissue [2–5, 11], which reduces the sensitivity of detection in copy number aberrations and confounds the transcriptomic profiles due to contamination with diploid stromal and inflammatory cells.

The aims of this study were twofold: (1) to determine the genes whose expression levels correlate with gene copy numbers and the genes which are overexpressed when amplified in grade III (GIII) breast cancers; and (2) to determine canonical pathways and networks significantly enriched for these genes. We microdissected and analysed a series of 48 GIII invasive ductal carcinomas of no special type (IDC-NSTs) with high resolution microarray-based comparative genomic hybridisation (aCGH) and mRNA expression arrays. Our study focused on GIII IDC-NSTs only to avoid bias introduced by distinct patterns of genetic aberrations between grade I and III cancers [12] and between invasive ductal carcinomas and special types [14–16].

Materials and methods

Sample selection

A series of 64 consecutive GIII IDC-NSTs cases were retrieved from University Hospital La Paz, Madrid, Spain. This project has been approved by the local Research Ethics Committees of the authors' institutions.

Four-μm thick sections of the tumours were subjected to immunohistochemistry with antibodies against oestrogen receptor, (ER, ID5, 1:30, Novocastra), progesterone receptor (PR, 1A6, 1:30, Novocastra), cytokeratin (CK) 5/6 (D5/16 B4, 1:25, DAKO) and epidermal growth factor receptor (EGFR) (EGFR, 31G7, 1:50, Zymed) essentially as previously described [1]. Immunohistochemistry for HER2 was performed using the Herceptest kit (Dako, Glostrup, Denmark). HER2 expression was scored according to the ASCO/CAP guidelines [17]; only cases displaying a 3+ score were considered positive, cases displaying a 2+ score were labelled as equivocal. In equivocal cases, *HER2* gene status was determined using the FDA approved SpotLight chromogenic in situ hybridisation (CISH) *HER2* probes. All markers were scored by two pathologists (SMP-R and JSR-F), blinded to the results of the aCGH, expression array and CISH analysis. Tumours were classified into HER2, luminal and basal-like subgroups utilising the criteria of Nielsen et al. [18].

Nucleic acid extraction

Ten consecutive 8-μm thick, frozen sections of 64 GIII tumours were microdissected as previously described [14, 19] to obtain a percentage of tumour cells in the remaining tissue greater than 90%. DNA was extracted as previously described [12]. DNA concentration was measured with Picogreen® according to the manufacturer's instructions (Invitrogen, Paisley, UK). RNA was extracted using Trizol according to the manufacturers' instructions (Invitrogen, UK) and quantified using the Agilent 2100 Bioanalyzer with RNA Nano LabChip Kits (Agilent Biosystems). In 16 cases, DNA and/or RNA were of insufficient quantity for aCGH and/or expression array analysis and were excluded from the study. The final dataset comprises 48 cases, whose clinico-pathological characteristics are summarised in Supplementary Table 1.

Microarray-based comparative genomic hybridisation

The 32K bacterial artificial chromosome (BAC) re-array collection (CHORI) tiling path aCGH platform used for this study was constructed in the Breakthrough Breast Cancer Research Centre [14]. DNA labelling, array hybridisations and image acquisition were performed as described in Natrajan et al. [1].

Cases with >10% of clones missing and clones where data were not available in ≥10% of cases were excluded. Log2 ratios were normalised for spatial and intensity-dependent biases using a two-dimensional LOESS regression followed by a BAC-dependent bias correction as

previously described [14, 20]. Polymorphic BACs identified in an analysis of 50 male:female and female:female hybridisations were filtered out. This left a final dataset of 31,544 clones with unambiguous mapping information according to the March 2006 build (hg18) of the human genome (http://www.ensembl.org). Data were smoothed using the circular binary segmentation (cbs) algorithm [6]. A categorical analysis was applied to the BACs after classifying them as representing gain, loss or no-change according to their smoothed Log2 ratio values [21]. Threshold values were chosen to correspond to three standard deviations of the normal ratios obtained from the filtered clones mapping to chromosomes 1-22, assessed in multiple hybridisations between DNA extracted from a pool of male and female blood donors as previously described [21, 22] (Log2 ratio of ± 0.08). Low level gain was defined as a smoothed Log2 ratio of between 0.08 and 0.45, corresponding to approximately 3–5 copies of the locus, whilst gene amplification was defined as having a Log2 ratio >0.45, corresponding to more than 5 copies [19].

Gene expression profiling

Gene expression profiling was performed using the Illumina human WG6 version 2 expression array according to the manufacturer's protocol. Briefly, 200 ng of each RNA sample was amplified using the Illumina Totalprep RNA amplification kit following the manufacturer's instructions (Ambion, UK). Labelled cRNA was subsequently hybridised to Illumina Sentrix-human 6 version 2 expression bead-chips. Full details on RNA amplification and hybridisation can be found at www.illumina.com. Raw gene expression values were robust-spline normalised using the Bioconductor lumi package (http://www.bioconductor.org/packages/2.3/bioc/html/lumi.html) in R. Genes were mapped to their genomic location using the lumiHumanAllv2 annotation database available from Bioconductor. Only Illumina transcript probes with detection P values <0.01 in $>25\%$ of samples were included; this resulted in a dataset of 12,699 transcriptionally regulated probes with accurate and unequivocal mapping information. Gene expression data are publicly available at ArrayExpress http://www.ebi.ac.uk/microarray-as/ae/ (accession number: E-TABM-543).

Integrated aCGH and expression analysis

To identify genes whose expression correlated with genetic copy number changes, cbs-smoothed aCGH data were used to assign the median aCGH states for each of the 12,699 genes in the gene expression dataset using the median values for all BACs that overlap with the genomic positions of each gene. This resulted in a 1:1 matrix of expression data and aCGH values used in correlations. Pearson's correlations were performed between mRNA expression log intensity values and median cbs-smoothed ratios derived from aCGH analysis for each gene. P values for each test were adjusted with Benjamini and Hochberg multiple P value adjustment [6, 23].

To identify genes consistently overexpressed when amplified in regions of recurrent amplification, Mann–Whitney U tests were performed and adjusted with Benjamini and Hochberg multiple P value adjustment. An adjusted P value of <0.05 was considered significant.

For unsupervised analysis, we performed hierarchical clustering of the 48 primary tumours using 848 unique genes identified from the 1,300 intrinsic genes, as described by Hu et al. [24]. Genes were median centred, and average linkage clustering was performed using Pearson's correlation as the similarity metric on Cluster 3 [25]. Results were subsequently visualised using Java Treeview (http://jtreeview.sourceforge.net/). SAM software [26] was employed to determine the genes significantly differentially expressed between basal-like, HER2 and luminal tumours, adopting a false discovery rate of 3% after 1,000 permutations.

Ingenuity pathway analysis

To determine pathways and networks that were significantly regulated in the gene expression data of basal-like, HER2 and luminal GIII tumours, we performed pathway analysis using the Ingenuity Pathway Analysis (IPA) program (http://www.ingenuity.com). HUGO gene identifiers were mapped to networks available in the Ingenuity database and ranked by score. The score indicates the likelihood of the genes in a network being found together due to random chance. Using a 99% confidence level, scores of ≥ 3 are considered significant. IPA was performed essentially as previously described [16, 27] with the Gene-sets described in Fig. 1.

Exact hypergeometric probability analysis for gene list enrichment

Analysis for gene list enrichment was performed using hypergeometric probability analysis, whereby the number of genes in common between two groups was identified and a representation factor (the number of overlapping genes divided by the expected number of overlapping genes drawn from two independent groups) was calculated. A representation factor >1 indicates more overlap than expected of two independent groups, whereas a representation factor <1 indicates less overlap than expected. The probability of finding an overlap of that number of genes was then calculated using the hypergeometric probability formula: $C\ (D,\ x) \times C\ (N - D,\ n - x)/C\ (N,\ n)$. http://elegans.uky.edu/MA/progs/representation.stats.html.

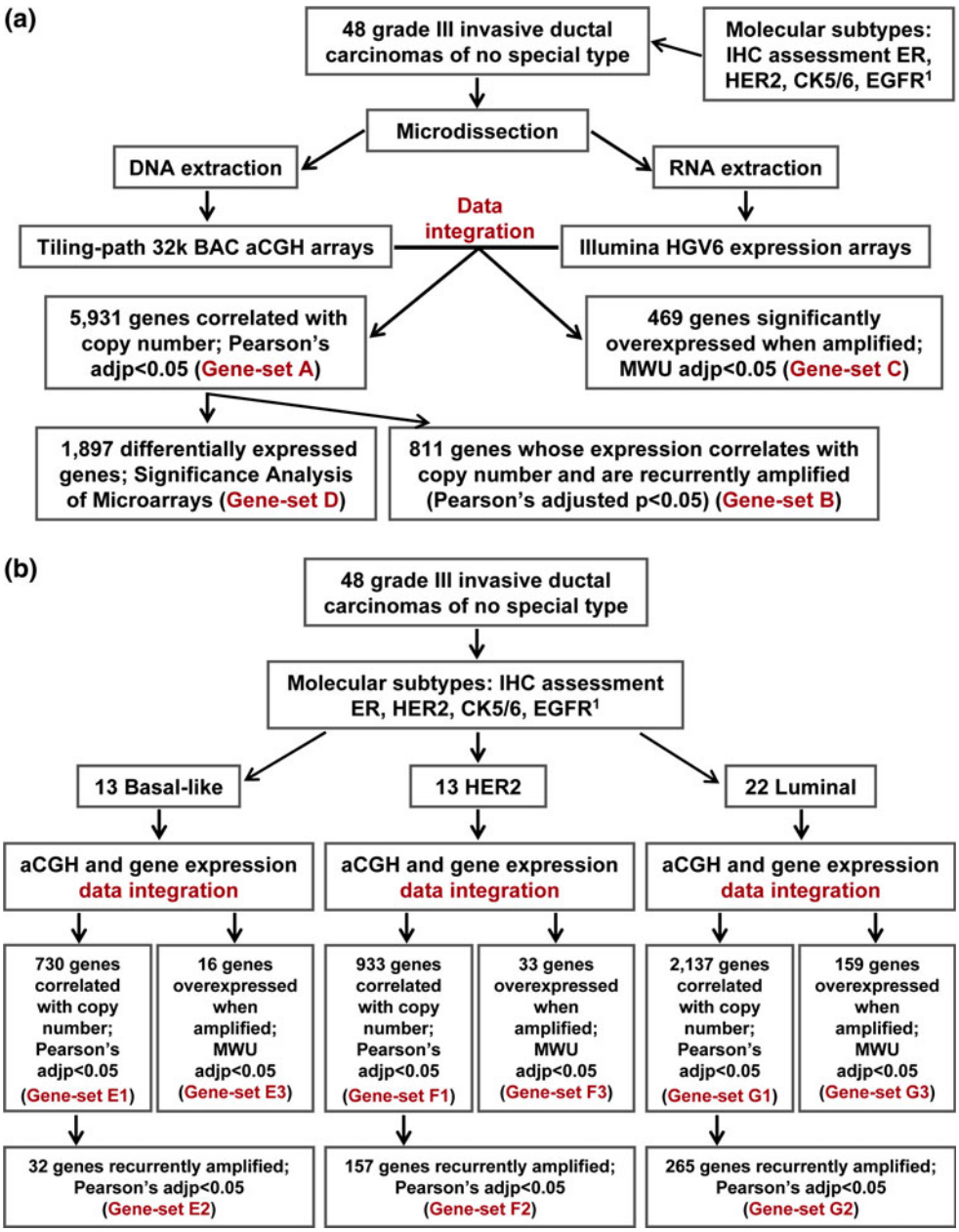

Fig. 1 Work flow diagram of integration of aCGH and gene expression data. **a** Analysis of cohort of 48 grade III invasive ductal carcinomas of no special type. **b** Analysis of 48 grade III invasive ductal carcinomas of no special type assigned to the molecular subtypes basal-like ($n = 13$), HER2 ($n = 13$) and luminal ($n = 22$) separately. [1]Nielsen et al. [18], *MWU* Mann–Whitney *U* test, *adjp* adjusted *P* value

Validation of aCGH and microarray-based expression profiling

All cases were reviewed by two pathologists (SMP-R and JSR-F), and representative areas of the 48 cases were included in a tissue microarray containing two 1-mm replicate cores from each case. These cases were used to validate selected amplifications found in aCGH.

Immunohistochemistry for cyclin D1 was performed with the SP4 rabbit monoclonal antibody (Neomarkers, Suffolk, UK) at 1:50 dilution as previously described [28] and scored using the Allred scoring system [28]. The immunohistochemical analysis was performed by two of the authors (SMP-R and JSR-F) with observers blinded to the results of the aCGH, gene expression and CISH analyses.

Chromogenic in situ hybridisation was performed using the Zymed/Invitrogen SpotLight CISH probes against *HER2*, *CCND1* and *EGFR* [21, 28, 29], and an in-house generated *PPM1D* probe [30]. Hybridisations and washes were performed as previously described [21, 28–30]. Gene amplification was defined as the presence of >5 signals, large gene clusters or a combination of signals and clusters in >50% of morphologically unequivocal neoplastic cells as previously described [21, 28–30]. CISH analysis was performed by two of the authors (SMP-R and FCG) with observers blinded to the results of the aCGH, gene expression and immunohistochemical analyses.

Quantitative real-time reverse transcriptase PCR

RNA from 48 microdissected cases was available for quantitative real-time reverse transcriptase PCR (qRT-PCR) analysis. First strand synthesis was performed as previously described [1], and qRT-PCR was performed with TaqMan chemistry on the ABI Prism 7900HT (Applied Biosystems), using the standard curve method [1, 21, 30]. Assays for PPM1D and TFRC were purchased from Applied Biosystems. PPM1D expression levels were normalised to gene TFRC (Assay on demand ID: Hs00174609_m1-TFRC and Hs00186230_m1-PPM1D) [30].

Results

Breast cancer molecular subtype classification

To determine the basal-like, luminal and HER2 molecular subtype classes, we subjected 48 GIII IDC-NSTs to a validated immunohistochemical surrogate of the molecular subtypes of breast cancer described by Nielsen et al. [18]. This taxonomy was employed given that it can be used to classify archival tumour samples and more accurately identifies the clinically relevant subgroup of HER2-positive/amplified tumours (i.e. amenable to therapy with anti-HER2 agents). Using Hu et al.'s [24] or Parker et al.'s [31] microarray-based classification, a proportion of ER-positive *HER2* amplified cancers are classified as of luminal B phenotype. Based on the expression of HER2, ER, CK 5/6 and EGFR, 13, 22 and 13 tumours were classified as HER2, luminal and basal-like phenotype, respectively (Supplementary Table 1). As expected, a good agreement between the classification of tumours into basal-like, HER2 and luminal using Nielsen et al.'s [18] immunohistochemical surrogate and by expression profiling using the Hu et al.'s centroids [24] was observed (Kappa score = 0.753, $P < 1 \times 10^{-11}$, Supplementary Table 1). As expected [31], cases harbouring *HER2* amplification were consistently classified as of HER2 phenotype by Nielsen et al. [18] and fell into either HER2 or luminal subgroup using the Hu et al.'s centroids [24].

Identification of genes whose expression is associated with gene copy number

To define genes whose expression is significantly correlated with gene copy number, we extracted DNA and RNA from microdissected samples and performed aCGH analysis, using a BAC array platform (actual resolution 50 kb) and genome-wide gene expression profiling using the Illumina WG6 v2 platform. We next overlaid the aCGH and gene expression data of our 48 GIII IDC-NSTs. Out of the 12,699 significantly regulated genes, 5,931 (46.70%) displayed mRNA expression levels that significantly correlated with copy number (Pearson's correlation adjusted $P < 0.05$, Supplementary Table 2, Gene-set A, Fig. 1a). This list included genes that have previously been found to have increased expression levels determined by gains in copy number, such as *ORAOV1*, *ASH2L*, *FADD*, *RPS6KB1*, *NME1*, *EIF4EBP2*, *GRB7*, *INTS2*, *PERLD1*, *MYST3*, *PIGS*, *INTS4*, *TLK2*, *PPM1D*, *PROSC*, *STARD3*, *ERBB2* and *CTTN* [3–6, 11].

To identify potential amplicon drivers, we determined genes whose expression correlates with copy number and are recurrently amplified ($n = 811$, Pearson's correlation adjusted $P < 0.05$; amplified in 2 or more tumours, Supplementary Table 3; Gene-set B) and genes significantly overexpressed when amplified ($n = 469$, Mann–Whitney U test adjusted $P < 0.05$, Supplementary Table 3; Gene-set C). Not surprisingly, these lists of genes included known drivers of amplicons on 1q, 8p, 8q, 11q13, 17q12-q21, 17q22-q25 and 20q13 [1, 3–5, 7, 11, 32–36]. This unbiased approach accurately identified known amplicon drivers, including *HER2*, *STARD3*, *GRB7* [37] on 17q12, *ORAOV1* [38] on 11q13 and *PPM1D* [1, 5, 30] on 17q23.2, genes whose expression and activity have been shown to be required for the proliferation and/or survival of cancer cells harbouring their amplification. In addition, this analysis has also revealed additional genes on chromosome 17q that do not pertain to the smallest region of amplification on 17q12-q21 or on 17q22-q25, but are significantly overexpressed when amplified (e.g. *MRPL45*, *FBXL20*, *MED1* and *ORMDL3*, Fig. 2a), suggesting that the reported complex patterns of genetic aberrations on chromosome 17 [39] may have a biological impact (i.e. overexpression of specific genes) and are not mere genomic 'noise'. Similar findings were observed in the genomic regions adjacent to the amplicons on 11q13.2-q13.3 and 17q21-q23 (Fig. 2b,c).

On the other hand, the 8q23.2-24.3 amplicon displayed a significantly more complex pattern of genetic aberrations without a well-defined core. Our unbiased analysis demonstrates that *UTP23*, *RAD21*, *MED30*, *MRPL13*, *MTBP*, *C8orf76*, *ZHX1*, *C8orf32*, *FBXO32*, *NDUFB9*, *KIAA0196*, *NSMCE2*, *DDEF1*, *EFR3A*, *PHF20L1*, *ST3GAL1* and *TRAPPC9* (Fig. 2d) are overexpressed when amplified and constitute potential drivers of this amplicon. It should be noted that *MYC*, which maps to this region, has expression levels that correlate with copy number, but is not consistently overexpressed when amplified, as previously described [40]. Taken together with the results of previous functional and correlative studies [3, 4, 7, 11, 13, 37, 38], our analysis seems to identify accurately genes whose expression correlates with copy number and potential amplicon drivers.

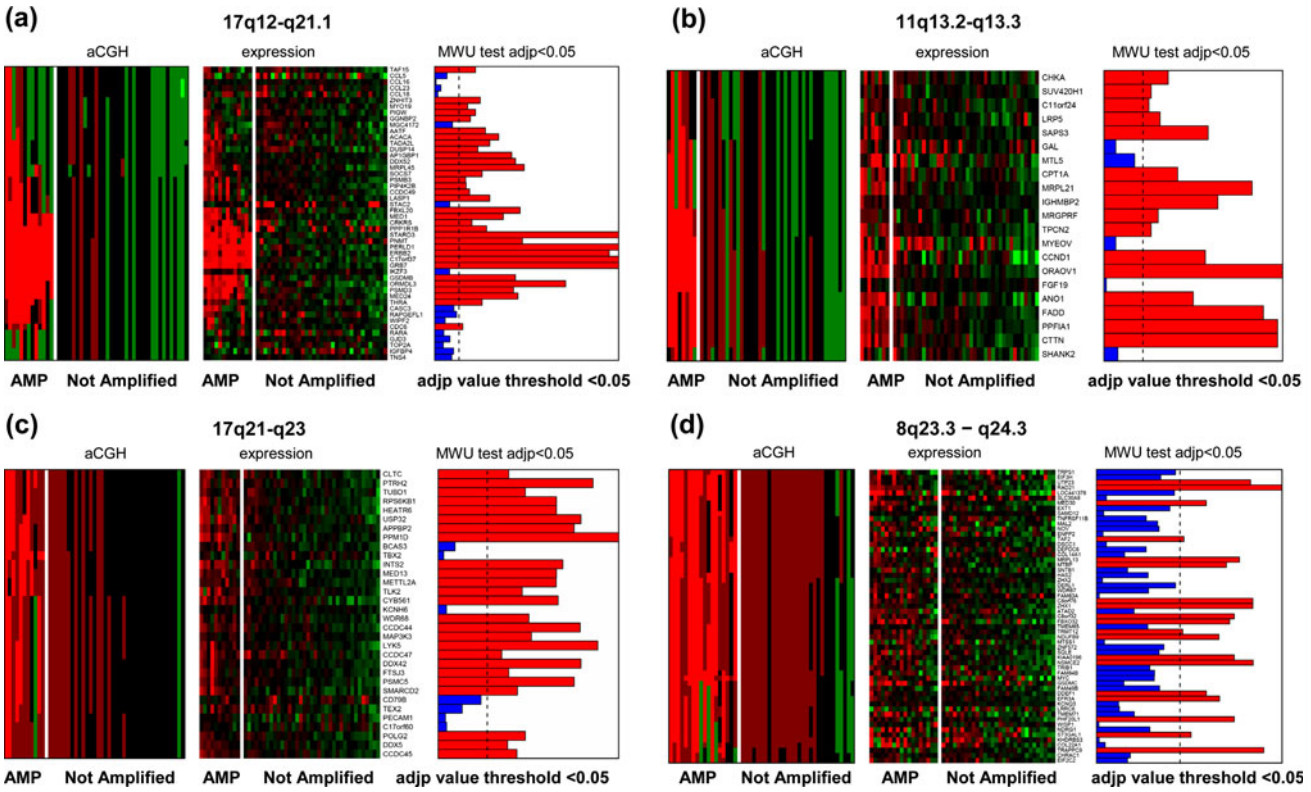

Fig. 2 Matched heatmaps of gene expression and aCGH within regions of recurrent amplification. **a** 17q12-q21.1, **b** 11q13.2-q13.3, **c** 17q21-q23 and **d** 8q23.3-q24.3. For each amplicon, genes within the amplified region are recovered and median aCGH values and states are assigned. Samples are separated into those harbouring amplification within the region and those that do not. Expression and aCGH values are depicted in two matching heatmaps (aCGH states on the *left* and expression values on the *right*) in which the genes are ordered according to their chromosomal position. *Bar plots* show the result of a Mann–Whitney *U* test for expression as a continuous variable and gene amplification as the grouping variable. *Bars* in *red* show adjusted *P* values of less than 0.05. aCGH: *green* copy number loss; *black* no copy number change; *dark red* copy number gain; *bright red* gene amplification. Gene expression: *green* downregulation, *red* upregulation; *MWU* Mann–Whitney *U* test, *adjp* adjusted *P* value

Molecular pathways regulated by gene copy number differ in basal-like, HER2 and luminal cancers

To determine which of the 5,931 genes whose expression levels are significantly correlated with gene copy number are associated with each molecular subtype of GIII breast cancers (i.e. basal-like, HER2 and luminal), multiclass significance analysis of microarray (SAM) was performed [26] using 1,000 permutations and accepting a false discovery rate of 3%. SAM analysis revealed that 1,897 genes (Geneset D; Fig. 1a) were differentially expressed between GIII IDC-NSTs of basal-like, HER2 and luminal molecular subtypes (Supplementary Table 4). As expected, genes pertaining to the *HER2* amplicon, *PERLD1*, *C17orf37*, *GRB7*, *STARD3* and *HER2* (*ERBB2*) were significantly overexpressed in cancers of HER2 phenotype and were the top ranking genes in the SAM analysis.

Ingenuity Pathway Analysis (IPA) of these 1,897 transcripts revealed networks and pathways enriched for genes whose expression correlates with copy number and are differentially expressed between molecular subtypes. These included networks related to DNA replication and repair, and cell cycle (Fig. 3; Supplementary Fig. 1; Supplementary Table 5). For instance, the network 'Gene Expression, DNA Replication, Recombination, and Repair, Cell Cycle, Cancer' (score 40, Fig. 3) encompasses *MYC* and other key genes such as *MDC1*, a mediator of DNA damage checkpoint; *MSH3*, *MLH1* and *MLH3*, which play a role in mismatch repair; and *RAD50*, which is involved in DNA double-strand break repair. Basal-like cancers displayed recurrent copy number gains/amplifications and upregulation of *MYC* and *MDC1* and losses and downregulation of *RAD50* and genes related to mismatch repair (i.e. *MSH3*, *MLH1* and *MLH3*) compared to HER2 and luminal cancers (Fig. 3). The network 'Gene Expression, DNA Replication, Recombination, and Repair, Cell Cycle' (score 40, Supplementary Fig. 1) was centred around key genes involved in cell cycle such as *CCND1*, *E2F3*, *CCNA2* and *CCNDBP1*, with basal-like tumours showing recurrent deletions and downregulation of *CCND1* and

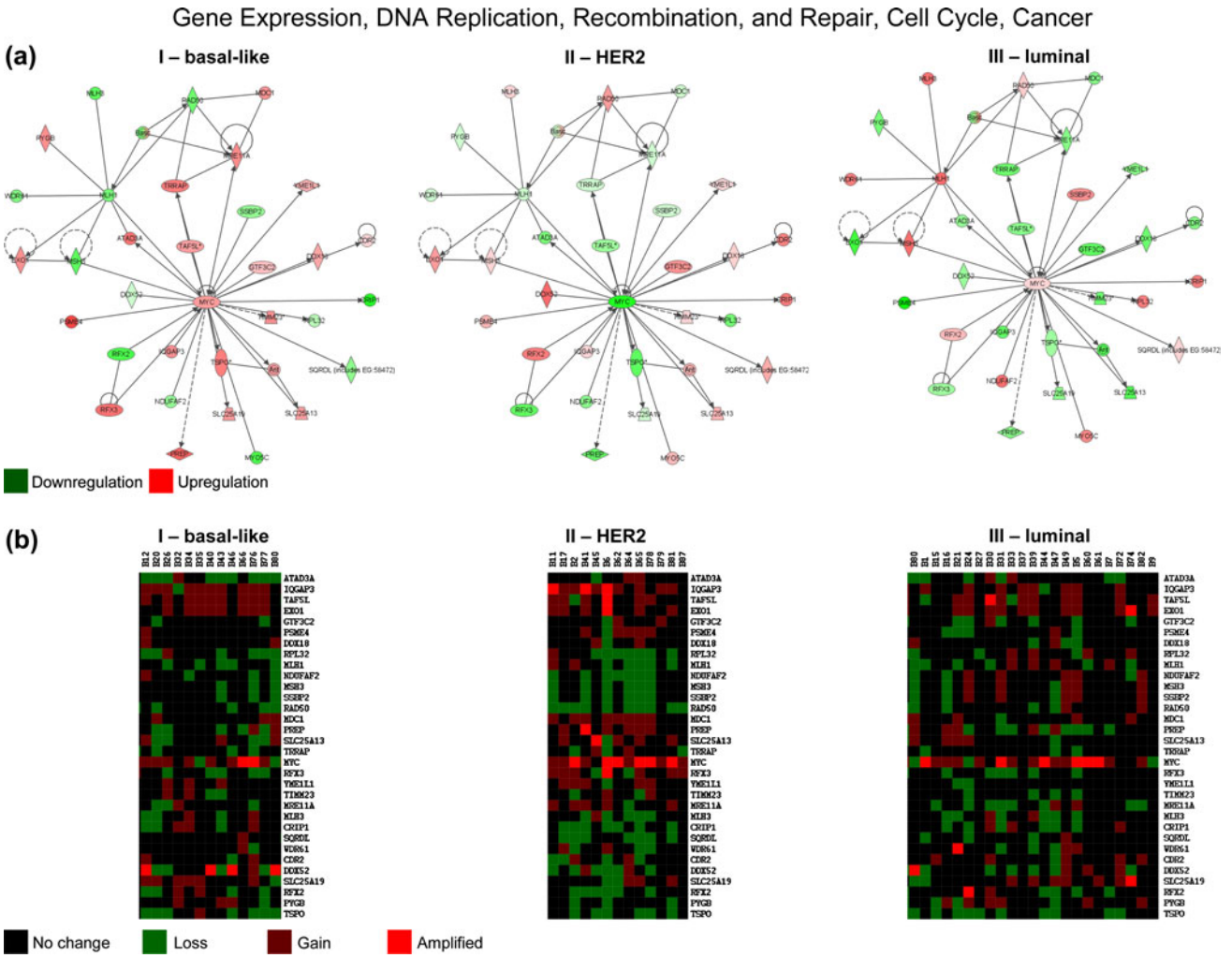

Fig. 3 Network enriched for genes whose expression correlates with copy number and are significantly differentially expressed between basal-like, HER2 and luminal GIII invasive ductal carcinomas of no special type, defined by Ingenuity Pathway Analysis software. **a** 'Gene Expression, DNA Replication, Recombination, and Repair, Cell Cycle, Cancer' network (score 40) in (I) basal-like, (II) HER2 and (III) luminal tumours. *Green* downregulation, *red* upregulation. **b** Matched heatmaps of aCGH categorical data for each of the genes differentially expressed in (I) basal-like, (II) HER2 and (III) luminal tumours. *Green* copy number loss; *black* no copy number change; *dark red* copy number gain; *bright red* gene amplification

copy number gains and upregulation of *E2F3* and HER2, whereas luminal cancers harboured recurrent amplifications and overexpression of *CCND1* and less frequent gains of *E2F3* (Supplementary Fig. 1).

In agreement with the above, IPA also identified the canonical pathways 'Cell Cycle: G1/S Checkpoint Regulation' and 'Role of BRCA1 in DNA Damage Response' as significantly enriched for genes whose expression correlates with copy number and are significantly differentially expressed in breast cancer molecular subtypes ($P < 0.001$ and $P = 0.0069$, respectively) (Supplementary Table 6). This analysis revealed that the expression levels of distinct components of the G1/S checkpoint regulation pathway are differentially expressed in basal-like, luminal and HER2 cancers due to specific patterns of copy number aberrations (Fig. 4). Recurrent gains/amplifications and upregulation of *CDK6*, *p16INK4A*, *E2F*, *CCNE1* and *MYC* were observed in basal-like breast cancers, whereas *CCNE1* and *CCND1* were recurrently gained/amplified and upregulated in HER2 tumours, and *CCND1* in luminal tumours. *SKP1* and *SKP2* were recurrently deleted and downregulated in basal-like breast cancers, whereas recurrent gains and overexpression of *SKP2* were found in HER2 cancers, and recurrent gains and overexpression of both genes were observed in luminal cancers. Interestingly, p16 (*CDKN2A*), a gene that has been reported to be overexpressed in the majority of basal-like breast cancers [41], was recurrently gained and overexpressed in these tumours; whereas HER2 and luminal cancers harboured recurrent deletions and downregulation of this gene. Genes involved in BRCA1 DNA damage control

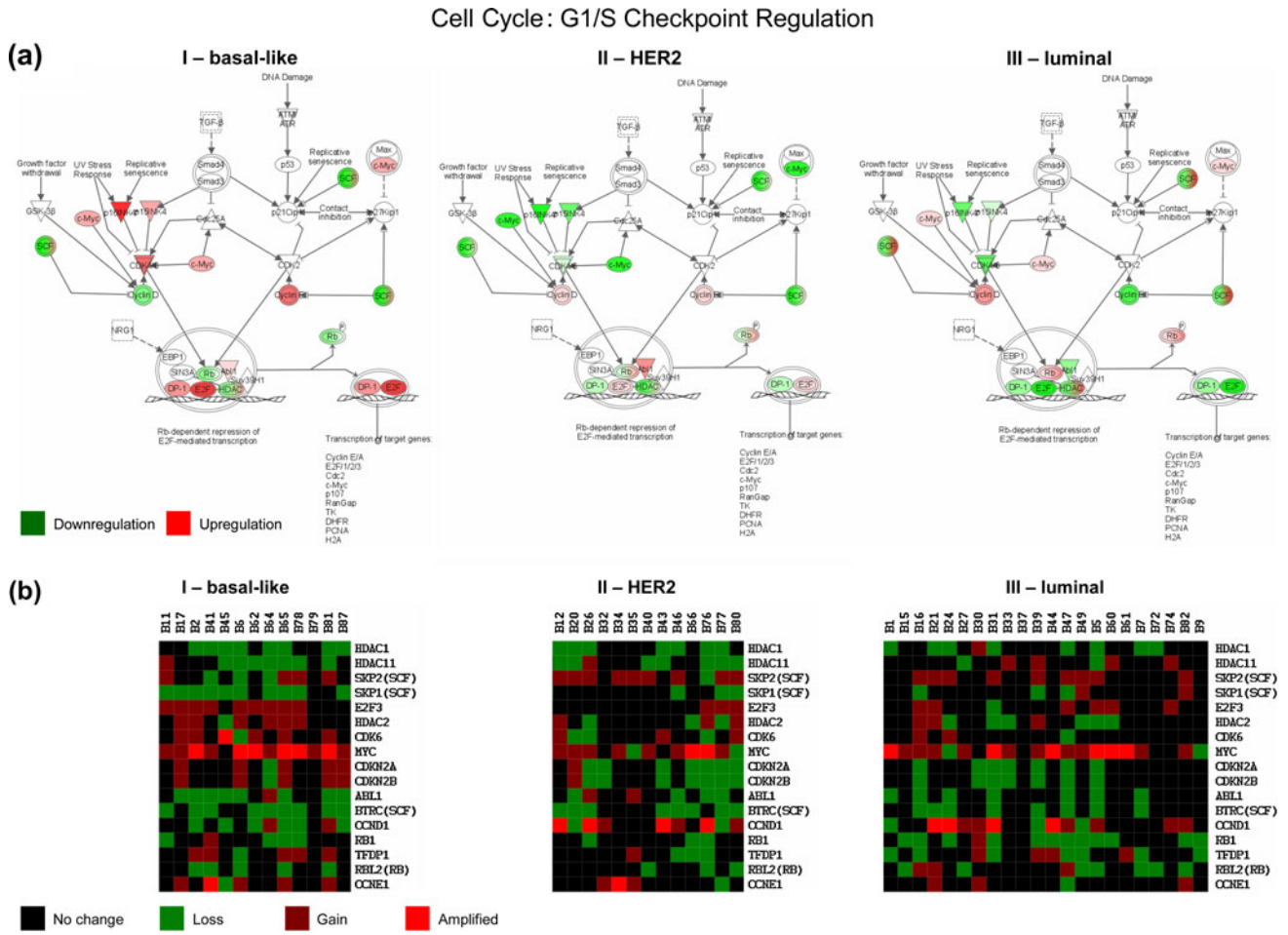

Fig. 4 Canonical pathway 'Cell Cycle: G1/S Checkpoint Regulation' significantly enriched for genes whose expression correlates with copy number and differentially expressed between basal-like, HER2 and luminal GIII tumours ($P < 0.001$), defined by Ingenuity Pathway Analysis software. **a** Canonical pathway in (I) basal-like, (II) HER2 and (III) luminal tumours; *green* downregulation, *red* upregulation. **b** Matched heatmaps of aCGH data of (I) basal-like, (II) HER2 and (III) luminal tumours. *Green* copy number loss; *black* no copy number change; *dark red* copy number gain; *bright red* gene amplification

and DNA repair mechanisms, including *RB*, *MLH1* and *RAD50*, were recurrently deleted and downregulated and *E2F3*, *MRE11A* and *FANCE* were recurrently copy number gained and upregulated in basal-like cancers compared with HER2 and luminal tumours (Supplementary Fig. 2). Although *BRCA1* did not reach significance in our multi-class SAM analysis, its expression is associated with copy number (Pearson's correlation adjusted $P = 0.0433$), and lower expression is associated with basal-like tumours ($P = 0.0283$, *t*-test, homoscedastic; adjusted $P = 0.01499$, Mann–Whitney *U* test). In addition, 'Protein Ubiquitination' ($P < 0.001$) and 'IGF-1' ($P = 0.0018$) were amongst the top canonical pathways significantly enriched for genes whose expression correlates with copy number (Supplementary Table 6).

These results suggest that the transcriptional regulation of molecular pathways may be determined by the pattern of genetic aberrations in breast cancers and that distinct pathways and networks are driven by specific patterns of copy number aberrations in basal-like, HER2 and luminal tumours. To corroborate the above hypotheses, we overlaid the aCGH and expression data for each subtype separately (Pearson's correlation adjusted $P < 0.05$, Supplementary Table 7). This resulted in a list of 730 (Gene-set E1), 933 (Gene-set F1) and 2,137 (Gene-set G1; see Fig. 1b) genes whose expression is associated with copy number in 13 basal-like, 13 HER2 and 22 luminal cancers, respectively. A significant overlap as defined by exact hypergeometric probability analysis was found for the genes identified as copy number regulated with both approaches in basal-like (Gene-set D vs. Gene-set E1, Representation factor: 2.0, $P < 10^{-25}$), HER2 (Gene-set D vs. Gene-set F1, Representation factor: 2.1, $P < 10^{-37}$) and luminal (Gene-set D vs. Gene-set G1, Representation factor: 1.9, $P < 10^{-71}$) cancers.

Ingenuity Pathway Analysis of Gene-sets E1, F1 and G1 (Pearson's correlation adjusted $P < 0.05$) revealed

networks and pathways independently associated with each molecular subtype (Supplementary Tables 6, 8). This analysis revealed a heterogeneous pattern of losses, gains and amplifications of genes whose expression levels are correlated with copy number between the tumour subtypes leading to the activation of distinct networks and pathways. In basal-like cancers (Gene-set E1), the 'Oestrogen Receptor Signalling' ($P = 0.00178$, Fig. 5) and 'Protein Ubiquitination' ($P = 0.00181$) canonical pathways were significantly enriched for genes whose expression correlates with copy number (Supplementary Table 6). Of these genes pertaining to the 'Oestrogen Receptor Signalling' pathway, the vast majority were lost and downregulated, with the exception of *PgR*, which was gained in 3 and amplified in one basal-like case (B11, which expressed high mRNA and protein levels of progesterone receptor (PR), Supplementary Table 1). A strong correlation was observed between ER and PR status as defined by immunohistochemistry and ESR1 and PgR mRNA expression levels as defined by microarray-based expression analysis (Mann–Whitney *U* test, $P = 5.5 \times 10^{-9}$ and $P = 1.2 \times 10^{-7}$, respectively; Supplementary Table 1). The 'Oestrogen Receptor Signalling' canonical pathway was also significantly enriched for genes whose expression correlates with copy number in HER2 cancers (Gene-set F1, $P = 0.01288$, Fig. 5); however, the genes that were copy number regulated in each subtype were largely distinct. In HER2 tumours, *RAF1* and *ERK1/2* were recurrently lost, whereas gains of multiple components of the RNA polymerase II complex, including *POLR2K* and TATA-binding protein-associated factor genes (i.e. *TAF4* and *TAF15*), were observed. In HER2 cancers, the 'Nucleotide Excision Repair Pathway' was significantly enriched for genes whose expression correlates with copy number ($P = 0.0214$), with recurrent gains and amplifications of *POLR2K* and losses of *RPA2* (Fig. 5; Supplementary Table 6).

These results suggest that several of the molecular pathways and networks specifically associated with basal-like, HER2 and luminal cancers are driven, at least in part, by specific genetic aberrations. Importantly, DNA copy number changes appear to be involved in the downregulation of oestrogen receptor signalling, one of the hallmark features of basal-like cancers [42, 43], and in the dysfunctional status of gene pathways and networks related to various types of DNA repair mechanisms.

Pathway analysis identifies novel potential therapeutic targets in GIII IDC-NSTs of basal-like, HER2 and luminal molecular subtype

Previous studies have demonstrated that genes whose expression correlates with copy number and are recurrently amplified and genes consistently overexpressed when amplified are likely to be amplicon drivers and may constitute therapeutic targets [1, 3, 5, 30, 36, 44]. To identify potential therapeutic targets specific for subgroups of basal-like, HER2 and luminal cancers, we defined (1) genes whose expression correlates with copy number in each molecular subtype individually (Pearson's correlation adjusted $P < 0.05$) and are recurrently amplified in each group and (2) genes overexpressed when amplified in each group individually (Mann–Whitney *U* adjusted $P < 0.05$) (Fig. 1b). The first analysis resulted in a list of 32, 157 and 265 genes in basal-like, HER2 and luminal tumours, respectively (Gene-sets E2, F2 and G2, respectively), whereas the second resulted in 16, 33 and 159 genes in basal-like, HER2 and luminal tumours, respectively (Gene-sets E3, F3 and G3, respectively), (Supplementary Tables 9, 10). A significant overlap as defined by exact hypergeometric probability analysis was found between the genes identified as recurrently amplified and genes overexpressed when amplified in basal-like (Gene-set E2 vs. Gene-set E3, Representation factor: 372.0, $P < 10^{-40}$), HER2 (Gene-set F2 vs. Gene-set F3, Representation factor: 68.8, $P < 10^{-50}$) and luminal (Gene-set G2 vs. Gene-set G3, Representation factor: 44.3, $P < 10^{-252}$) tumours.

Genes whose expression correlates with copy number and are recurrently amplified in basal-like cancers (Gene-set E2) mapped to 1q22-q24, 1q44, 8q24.1, 10p15 and 19q12-q13. IPA revealed that the networks 'Gene Expression, Cancer, Gastrointestinal Disease' (score 46) and 'Gene Expression, Cancer, Cell Death' (score 17) were significantly enriched for these genes (Supplementary Table 11; Supplementary Fig. 3). Interestingly, different components of these networks were amplified in individual cases. Given the limited sample size, only genes mapping to 19q12-q13 were significantly overexpressed when amplified in basal-like cancers after multiple comparison adjustment (Gene-set E3; Fig. 1b). These genes, including *C19orf2*, *MED30*, *SUPT 5H*, *SERTAD1* and *SERTAD3*, were part of the network 'Cancer, Cell Death, Endocrine System Disorders' (score 40, Supplementary Fig. 3). Notably, the p21-activated kinase 4 (*PAK4*) was present in both 'Gene expression, Cancer, Cell Death' and 'Cancer, Cell Death, Endocrine System Disorders' networks and has recently emerged as a potential therapeutic target [45].

In HER2 cancers, the network 'Lipid Metabolism, Molecular Transport, Nucleic Acid Metabolism' (score 37) was significantly enriched for recurrently amplified genes whose expression correlates with copy number (Gene-set F2) and the network 'Amino Acid Metabolism, Post-Translational Modification, Small Molecule Biochemistry' (score 29, Fig. 6; Supplementary Tables 11, 12) was significantly enriched for genes overexpressed when amplified (Gene-set F3; Fig. 1b). These networks were composed of genes mapping not only to the *HER2* amplicon (17q12-

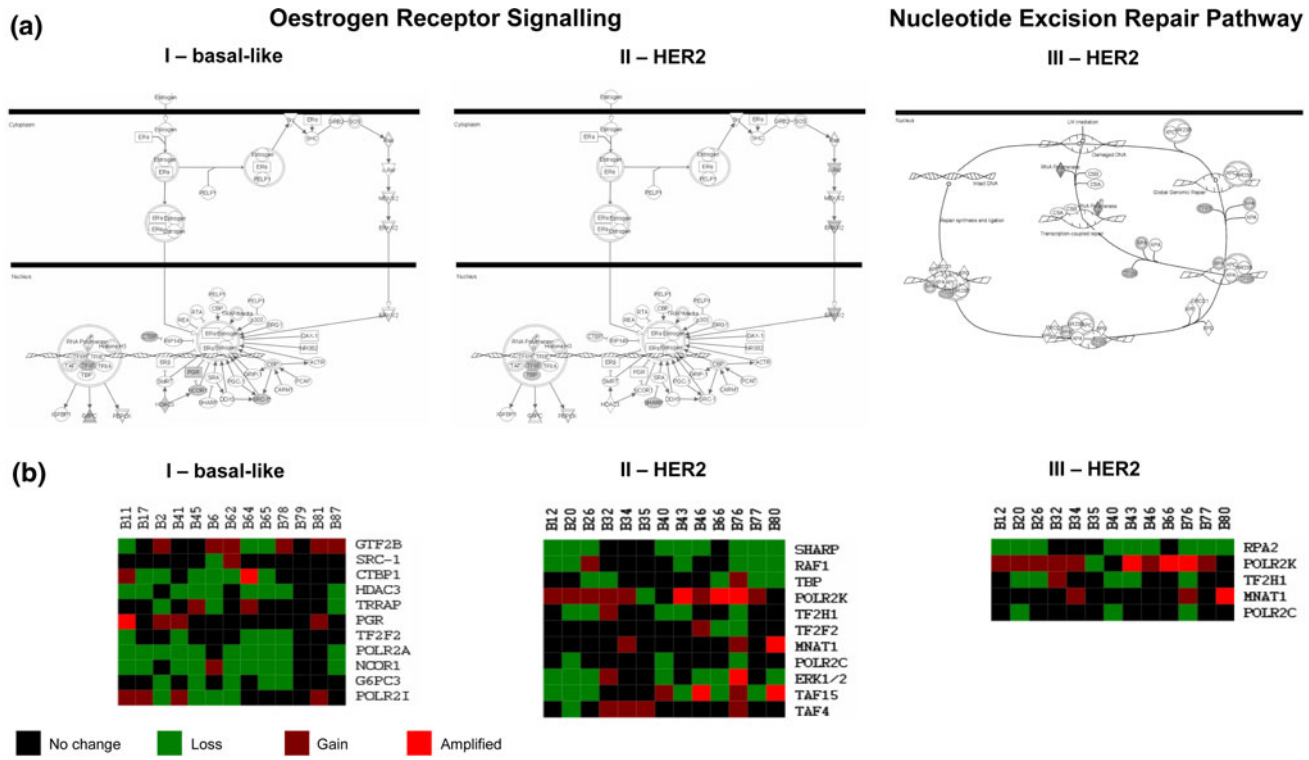

Fig. 5 Canonical pathways enriched for genes whose expression correlates with copy number in 13 basal-like and 13 HER2 GIII tumours, defined by Ingenuity Pathway Analysis software. **a** Canonical pathways (I) 'Oestrogen Receptor Signalling' ($P = 0.0018$) in basal-like, (II) 'Oestrogen Receptor Signalling' ($P = 0.013$) in HER2 and (III) 'Nucleotide Excision Repair Pathway' ($P = 0.0214$) in HER2 GIII tumours. Note that the expression levels of genes significantly correlate with gene copy number. **b** Matched heatmaps of aCGH data of (I) basal-like, (II) HER2 and (III) HER2 tumours. *Green* copy number loss; *black* no copy number change; *dark red* copy number gain; *bright red* gene amplification

q21), but also to distinct amplicons, including 8p12, 8q22-q24, 11q13-q14, 17p11 and 17q23, which were amplified in multiple combinations in each tumour (Fig. 6; Supplementary Tables 9, 10). Interestingly, *PPM1D*, a negative regulator of the p53 pathway and potential therapeutic target for a subgroup of patients with *HER2* amplified breast cancers [1], was present in both networks.

In luminal tumours, the 'ERK/MAPK Signalling' ($P = 0.00631$ and $P = 0.00978$, respectively) and 'PI3K/AKT Signalling' ($P = 0.01445$ and $P = 0.01047$, respectively) canonical pathways were significantly enriched for recurrently amplified genes whose expression correlates with copy number (Gene-set G2) and for genes overexpressed when amplified (Gene-set G3) (e.g., *PAK1*, *PRKAR1A*, *RPS6KB1*, *PPP2R5A*) (Supplementary Table 6; Fig. 7; Supplementary Fig. 4). Protein kinases of these two canonical pathways may constitute targets for therapeutic intervention by PI3K and/or MEK inhibitors [13, 46, 47]. Luminal breast cancers were also significantly enriched for genes in the networks 'Cell Cycle, Reproductive System Development and Function, Haematological System Development and Function' (score 40) and 'Reproductive System Disease, Cellular Assembly and Organisation, Cellular Movement' (score 33) (Supplementary Tables 11, 12; Fig. 7; Supplementary Fig. 4). Of note, *PPM1D* was part of the cell cycle network, providing additional evidence to our previous findings that *PPM1D* may also be a potential therapeutic target in a subgroup luminal breast cancers harbouring amplification of this gene [1].

Validation of aCGH and microarray-based expression profiling analysis

Two tissue microarrays containing the 48 tumours in our study were subjected to CISH with probes for *CCND1*, *EGFR*, *PPM1D* and *HER2*. A perfect agreement between aCGH and CISH results for *EGFR*, *PPM1D* and *HER2* was observed (Supplementary Table 1; Supplementary Fig. 5), whereas for *CCND1*, all but two cases showed concordant results (unweighted Kappa score = 0.86297; 95% CI 0.7099–1.0000).

A good correlation was observed between the microarray and qRT-PCR–defined expression levels of PPM1D (Pearson's r value = 0.81421, $P = 1.8 \times 10^{-11}$). In agreement with the results of the aCGH and expression array analysis, CISH and qRT-PCR-based assessment of *PPM1D* copy numbers and mRNA expression also

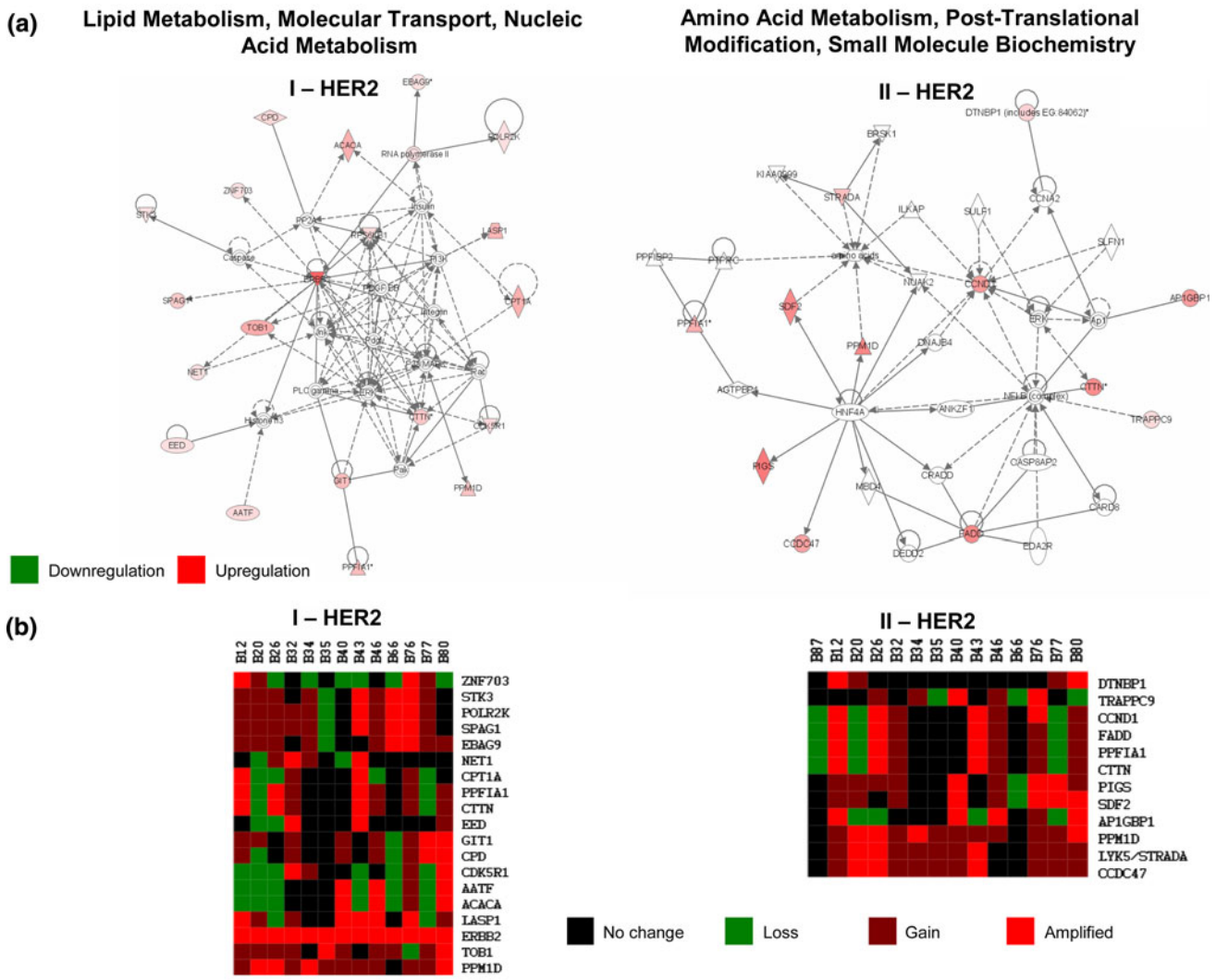

Fig. 6 Top ranked networks in 13 HER2 tumours enriched for genes whose expression correlates with copy number and are recurrently amplified (Pearson's correlation) and genes that are overexpressed when amplified (Mann–Whitney *U* test), defined by Ingenuity Pathway Analysis. **a** Top ranked networks in HER2 tumours (I) 'Lipid Metabolism, Molecular Transport, Nucleic Acid Metabolism' (Pearson's correlation) (score 37), (II) 'Amino Acid Metabolism, Post-Translational Modification, Small Molecule Biochemistry' (Mann–Whitney *U* test) (score 29); *Green* downregulation, *Red* upregulation. **b** Matched heatmap of aCGH data of HER2 tumours. *Green* copy number loss; *black* no copy number change; *dark red* copy number gain; *red* gene amplification

demonstrated that *PPM1D* is overexpressed when amplified (Mann–Whitney U test $= 5.0$, $P = 1.4 \times 10^{-5}$).

Immunohistochemical analysis of cyclin D1 using a semiquantitative scoring system showed a good correlation with CCND1 mRNA levels as defined by microarray analysis (Spearman's correlation coefficient $= 0.68943$, $P < 1 \times 10^{-6}$). Despite the impact of ER expression on cyclin D1 expression levels [28], CISH and immunohistochemistry based assessment of *CCND1* copy numbers and cyclin D1 protein expression, respectively, confirmed that *CCND1* is overexpressed when amplified (Mann–Whitney U test $= 62.0$, $P = 0.00694$), in agreement with aCGH and expression array analysis. Our results provide direct evidence for the accuracy of the aCGH and expression profiling methods employed in this study.

Discussion

The actual impact of gene copy number changes on the transcriptomic profiles of breast cancer molecular subtypes has been a matter of contention. Whilst HER2 tumours by definition are biologically driven by amplification and overexpression of *ERBB2*, the impact of genomic alterations on the transcriptome and phenotype of basal-like and luminal tumours is less clear [3]. Here, we show that

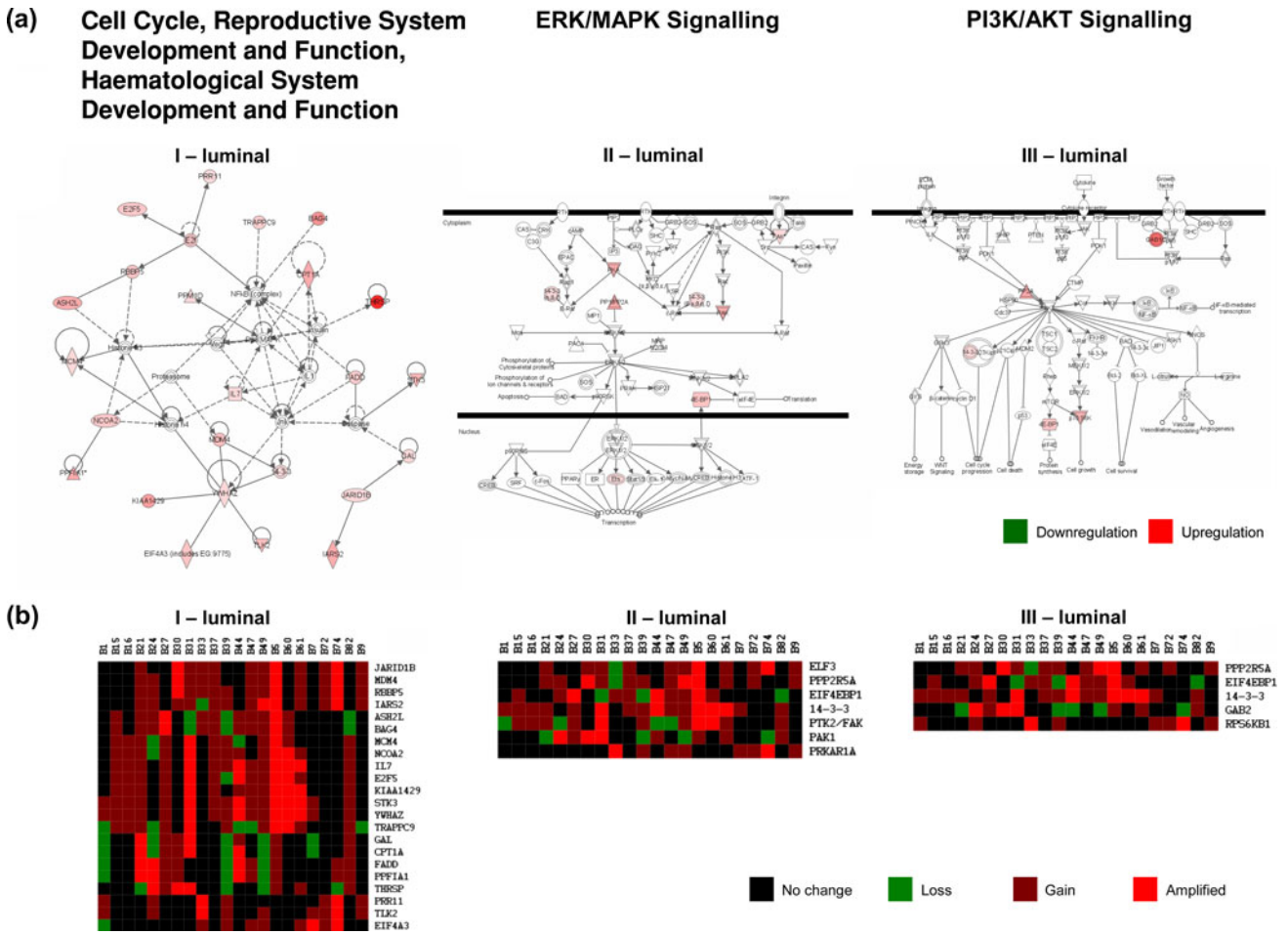

Fig. 7 Top ranking networks and canonical pathways identified by Ingenuity Pathway Analysis in 22 luminal tumours enriched for genes whose expression correlates with copy number and are recurrently amplified (Pearson's correlation). **a** Top ranked network and canonical pathways in luminal tumours (I) 'Cell Cycle, Reproductive System Development and Function, Haematological System Development and Function' (score = 40), (II) Canonical pathway 'ERK/MAPK Signalling' ($P = 0.0063$), (III) Canonical pathway 'PI3K/AKT Signalling' ($P = 0.0145$); *Green* downregulation, *Red* upregulation. **b** Matched heatmap of aCGH data of HER2 tumours. *Green* copy number loss; *black* no copy number change; *dark red* copy number gain; *red* gene amplification

distinct gene networks and canonical pathways are enriched for genes whose expression levels correlate with copy number in basal-like, HER2 and luminal tumours, providing evidence that differences between these molecular subtypes may be determined, at least in part, by their patterns of genetic aberrations.

Meta-analyses of microarray expression data have demonstrated that the main separation between molecular subtypes is driven by ER, HER2 and proliferation and that proliferation is a major determinant of breast cancer clinical behaviour [48, 49]. Our analysis demonstrates that networks and canonical pathways directly related to ER signalling (Fig. 5), HER2 signalling (Fig. 6) and proliferation (Fig. 4) are enriched for genes whose expression correlates with copy number. It should be noted that although ER pathway was enriched for genes whose expression correlates with copy number, *ESR1* mRNA levels and DNA copy number failed to show any statistically significant correlation (Pearson's correlation adjusted $P = 0.5667$) in our study, and *ESR1* was only amplified in a single ER-positive case (1.8%). This concurs with previous reports suggesting that only a small minority of ER-positive cancers harbour *ESR1* gene amplification and that ER expression is not significantly associated with *ESR1* copy numbers [50–53].

Our group and others have demonstrated that tumours arising in patients harbouring *BRCA1* germline mutations predominantly display a basal-like phenotype and that a subgroup of sporadic basal-like breast cancers harbours a dysfunctional *BRCA1* pathway [54–57]. Given the rarity of *BRCA1* somatic mutations, several mechanisms of BRCA1 pathway dysfunction in sporadic breast cancers have been proposed [43, 56, 57], including BRCA1 gene promoter methylation and overexpression of *ID4*, a negative regulator of BRCA1 [57]. In this study, we provide evidence that the BRCA1 pathway may be dysfunctional in sporadic

basal-like cancers due to gene copy number aberrations affecting genes pertaining to this pathway. The 'BRCA1 DNA Damage Response' pathway was enriched for genes whose expression correlates with copy number and are differentially expressed between the molecular subtypes (Supplementary Fig. 2). *BRCA1* expression itself was found to be associated with copy number and downregulated in basal-like tumours, suggesting a role for homologous recombination DNA repair defects in this subgroup of sporadic breast cancer. In addition, we provide evidence to suggest that overexpression of the BRCA1 negative regulator *ID4* in basal-like cancers is driven by copy number gains, as (1) the expression levels of this gene correlated with copy number in the whole cohort (Pearson's correlation adjusted $P < 0.01669$, Supplementary Table 2), (2) ID4 was expressed at significantly higher levels in basal-like breast cancer (Supplementary Table 4), and (3) *ID4* copy number gains were significantly more frequent in basal-like than in non–basal-like tumours (9/13 basal-like cancers and 10/35 non–basal-like cancers; Fisher's exact test, $P = 0.0187$). Taken together, our results support the contention that BRCA1 pathway dysfunction in sporadic basal-like breast cancers may be driven by genetic aberrations. Given the reported sensitivity of tumours with BRCA1 loss of function to PARP inhibitors, our findings provide further credence to the molecular basis for its potential efficacy in sporadic basal-like breast cancers [58].

By comparing our list of copy number regulated genes with that of published 'intrinsic gene' lists, we observed that 61/207 (29%) intrinsic genes described by Perou et al. [8] were present in our dataset, 166/365 (45.5%) intrinsic genes by Sorlie et al. [59], 435/929 (46.8%) unique identifiers of the 'Intrinsic/UNC' gene list by Hu et al. [24] and 684/1,488 (46%) intrinsic genes by Parker et al. [31] were significantly associated with gene copy number. The substantial enrichment of the published 'intrinsic gene' lists with genes whose expression correlates with copy number further emphasises the potential importance of genetic alterations in the molecular classification of breast cancers. Furthermore, it suggests that the differential expression of the 'intrinsic genes' in the distinct molecular subtypes may not only stem from their cell of origin [8, 9], but also as a result of genetic instability and subsequent acquisition of stochastic genetic aberrations.

To identify possible therapeutic targets, we used our integrative approach to distinguish recurrently amplified genes whose expression is associated with copy number and also genes that are consistently overexpressed when amplified. Our genome-wide, unbiased approach identified genes that have been functionally validated as potential therapeutic targets in specific breast cancer molecular subtypes (e.g. *HER2*, *PPM1D*, *PAK1* and *PPAPDC1B*—Supplementary Tables 9, 10) [1, 13, 44] and canonical pathways that have been exploited in preclinical therapeutic models for breast cancer (e.g. PI3K/AKT signalling in luminal cancers—Fig. 7 and Supplementary Fig. 4) [47], as well as a comprehensive list of genes that may constitute amplicon drivers and novel therapeutic targets.

Molecular subtype-specific analyses of genes consistently significantly overexpressed when amplified highlighted networks and canonical pathways enriched for potential therapeutic targets. The networks identified in basal-like tumours were found to be enriched for multiple genes involved in transcriptional regulation, suggesting that preclinical studies testing these transcriptional regulators as potential targets may be warranted. The p21-activated kinase 4 *PAK4*, which plays a role in integrin-mediated breast cancer cell motility [60], was significantly overexpressed when amplified in basal-like cancers. Novel strategies for targeting PAK4 and its family members in cancer are under investigation [45]. Ingenuity Pathway Analysis of HER2 cancers substantiates previous findings for the use of PPM1D-targeted therapies in this subgroup [1, 35] and revealed that PI3K and ERK/MAPK canonical pathways are significantly enriched for genes that are amplified and overexpressed in luminal tumours (e.g. *PAK1*, *PPP2R5A* and *PRKAR1A*). Interestingly, PAK1 overexpressing luminal breast cancer cell lines have recently been shown to be significantly more sensitive to MEK inhibition [13]. Moreover, the PI3K catalytic subunit inhibitor BEZ235 has been reported to cause selective cell growth inhibition in ER-positive breast cancer cell lines [47] and may constitute an interesting therapeutic approach for subgroups of patients with luminal breast cancer harbouring amplification of these genes.

In conclusion, high resolution integrative genomic analysis of basal-like, HER2 and luminal GIII breast cancers using microdissected tissues identified pathways and networks specific to each tumour type that are regulated by changes at the DNA level, and has identified key genes that govern distinct pathways. These results provide strong circumstantial evidence to suggest that the diversity of breast cancer and the molecular subtypes may stem, at least in part, from the distinct pattern of genetic aberrations found in these cancers. Furthermore, we have identified canonical pathways enriched for genes recurrently amplified or overexpressed when amplified in basal-like, HER2 and luminal cancers. Functional genomic studies to test the biological importance of these pathways and the therapeutic implication of inhibition of these pathways in appropriate preclinical models are warranted.

Acknowledgments This study was funded in part by Breakthrough Breast Cancer. B. Weigelt is supported by a Cancer Research UK Fellowship, J Palacios by RETICS: RD06/0020/0013 and PI 08/

6080971 from Instituto de Salud Carlos III and P07-CVI-03100 from Consejería de Innovación Junta de Andalucía.

References

1. Natrajan R, Lambros MB, Rodriguez-Pinilla SM et al (2009) Tiling path genomic profiling of grade 3 invasive ductal breast cancers. Clin Cancer Res 15:2711–2722
2. Bergamaschi A, Kim YH, Wang P et al (2006) Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 45:1033–1040
3. Chin K, DeVries S, Fridlyand J et al (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10:529–541
4. Chin SF, Teschendorff AE, Marioni JC et al (2007) High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 8:R215
5. Hu X, Stern HM, Ge L et al (2009) Genetic alterations and oncogenic pathways associated with breast cancer subtypes. Mol Cancer Res 7:511–522
6. Mackay A, Tamber N, Fenwick K, et al. (2009) A high-resolution integrated analysis of genetic and expression profiles of breast cancer cell lines. Breast Cancer Res Treat. doi:10.1007/s10549-008-0296-7
7. Neve RM, Chin K, Fridlyand J et al (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10:515–527
8. Perou CM, Sorlie T, Eisen MB et al (2000) Molecular portraits of human breast tumours. Nature 406:747–752
9. Sorlie T, Perou CM, Tibshirani R et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 98:10869–10874
10. Hicks J, Krasnitz A, Lakshmi B et al (2006) Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res 16:1465–1479
11. Adelaide J, Finetti P, Bekhouche I et al (2007) Integrated profiling of basal and luminal breast cancers. Cancer Res 67:11565–11575
12. Natrajan R, Lambros MB, Geyer FC et al (2009) Loss of 16q in high grade breast cancer is associated with estrogen receptor status: Evidence for progression in tumors with a luminal phenotype? Genes Chromosomes Cancer 48:351–365
13. Heiser LM, Wang NJ, Talcott CL et al (2009) Integrated analysis of breast cancer cell lines reveals unique signaling pathways. Genome Biol 10:R31
14. Marchio C, Iravani M, Natrajan R et al (2008) Genomic and immunophenotypical characterization of pure micropapillary carcinomas of the breast. J Pathol 215:398–410
15. Simpson PT, Reis-Filho JS, Lambros MB et al (2008) Molecular profiling pleomorphic lobular carcinomas of the breast: evidence for a common molecular genetic pathway with classic lobular carcinomas. J Pathol 215:231–244
16. Weigelt B, Horlings HM, Kreike B et al (2008) Refinement of breast cancer classification by molecular characterization of histological special types. J Pathol 216:141–150
17. Wolff AC, Hammond ME, Schwartz JN et al (2007) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 25:118–145
18. Nielsen TO, Hsu FD, Jensen K et al (2004) Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10:5367–5374
19. Marchio C, Natrajan R, Shiu KK et al (2008) The genomic profile of HER2-amplified breast cancers: the influence of ER status. J Pathol 216:399–407
20. Huang J, Gusnanto A, O'Sullivan K et al (2007) Robust smooth segmentation approach for array CGH data analysis. Bioinformatics 23:2463–2469
21. Arriola E, Marchio C, Tan DS et al (2008) Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines. Lab Invest 88:491–503
22. Reis-Filho JS, Simpson PT, Jones C et al (2005) Pleomorphic lobular carcinoma of the breast: role of comprehensive molecular pathology in characterization of an entity. J Pathol 207:1–13
23. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 57:289–300
24. Hu Z, Fan C, Oh DS et al (2006) The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96
25. Eisen MB, Spellman PT, Brown PO et al (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868
26. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98:5116–5121
27. Weigelt B, Kreike B, Reis-Filho JS (2008) Metaplastic breast carcinomas are basal-like breast cancers: a genomic profiling analysis. Breast Cancer Res Treat. doi:10.1007/s10549-008-0197-9
28. Reis-Filho JS, Savage K, Lambros MB et al (2006) Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic in situ hybridisation analysis. Mod Pathol 19:999–1009
29. Reis-Filho JS, Pinheiro C, Lambros MB et al (2006) EGFR amplification and lack of activating mutations in metaplastic breast carcinomas. J Pathol 209:445–453
30. Tan DS, Lambros MB, Rayter S et al (2009) PPM1D is a potential therapeutic target in ovarian clear cell carcinomas. Clin Cancer Res 15:2269–2280
31. Parker JS, Mullins M, Cheang MC et al (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27:1160–1167
32. Cheng KW, Lahad JP, Kuo WL et al (2004) The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. Nat Med 10:1251–1256
33. Fiddes RJ, Campbell DH, Janes PW et al (1998) Analysis of Grb7 recruitment by heregulin-activated erbB receptors reveals a novel target selectivity for erbB3. J Biol Chem 273:7717–7724
34. Hata T, Furukawa T, Sunamura M et al (2005) RNA interference targeting aurora kinase a suppresses tumor growth and enhances the taxane chemosensitivity in human pancreatic cancer cells. Cancer Res 65:2899–2905
35. Rayter S, Elliott R, Travers J et al (2008) A chemical inhibitor of PPM1D that selectively kills cells overexpressing PPM1D. Oncogene 27:1036–1044
36. Reis-Filho JS, Simpson PT, Turner NC et al (2006) FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas. Clin Cancer Res 12:6652–6662
37. Kao J, Pollack JR (2006) RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer 45:761–769
38. Jiang L, Zeng X, Yang H et al (2008) Oral cancer overexpressed 1 (ORAOV1): a regulator for the cell growth and tumor angiogenesis in oral squamous cell carcinoma. Int J Cancer 123:1779–1786
39. Orsetti B, Nugoli M, Cervera N et al (2004) Genomic and expression profiling of chromosome 17 in breast cancer reveals

complex patterns of alterations and novel candidate genes. Cancer Res 64:6453–6460
40. Guan Y, Kuo WL, Stilwell JL et al (2007) Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. Clin Cancer Res 13:5745–5755
41. Subhawong AP, Subhawong T, Nassar H et al (2009) Most basal-like breast carcinomas demonstrate the same Rb-/p16+ immunophenotype as the HPV-related poorly differentiated squamous cell carcinomas which they resemble morphologically. Am J Surg Pathol 33:163–175
42. Reis-Filho JS, Tutt AN (2008) Triple negative tumours: a critical review. Histopathology 52:108–118
43. Rakha EA, Reis-Filho JS, Ellis IO (2008) Basal-like breast cancer: a critical review. J Clin Oncol 26:2568–2581
44. Bernard-Pierrot I, Gruel N, Stransky N et al (2008) Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. Cancer Res 68:7165–7175
45. Eswaran J, Soundararajan M, Knapp S (2009) Targeting group II PAKs in cancer and metastasis. Cancer Metastasis Rev 28:209–217
46. Ihle NT, Powis G (2009) Take your PIK: phosphatidylinositol 3-kinase inhibitors race through the clinic and toward cancer therapy. Mol Cancer Ther 8:1–9
47. Crowder RJ, Phommaly C, Tao Y, et al. (2009) PIK3CA and PIK3CB inhibition produce synthetic lethality when combined with estrogen deprivation in estrogen receptor-positive breast cancer. Cancer Res 69:3955–3962
48. Desmedt C, Haibe-Kains B, Wirapati P et al (2008) Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res 14:5158–5165
49. Wirapati P, Sotiriou C, Kunkel S et al (2008) Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 10:R65
50. Reis-Filho JS, Drury S, Lambros MB et al (2008) ESR1 gene amplification in breast cancer: a common phenomenon? Nat Genet 40:809–810 author reply 10-2
51. Brown LA, Hoog J, Chin SF et al (2008) ESR1 gene amplification in breast cancer: a common phenomenon? Nat Genet 40:806–807 author reply 10-2
52. Horlings HM, Bergamaschi A, Nordgard SH et al (2008) ESR1 gene amplification in breast cancer: a common phenomenon? Nat Genet 40:807–808 author reply 10-2
53. Vincent-Salomon A, Gruel N, Lucchesi C et al (2007) Identification of typical medullary breast carcinoma as a genomic subgroup of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res 9:R24
54. Lakhani SR, Reis-Filho JS, Fulford L et al (2005) Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 11:5175–5180
55. Foulkes WD, Stefansson IM, Chappuis PO et al (2003) Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 95:1482–1485
56. Turner NC, Reis-Filho JS (2006) Basal-like breast cancer and the BRCA1 phenotype. Oncogene 25:5846–5853
57. Turner NC, Reis-Filho JS, Russell AM et al (2007) BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene 26: 2126–2132
58. Ashworth A (2008) A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. J Clin Oncol 26:3785–3790
59. Sorlie T, Tibshirani R, Parker J et al (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100:8418–8423
60. Zhang H, Li Z, Viklund EK et al (2002) P21-activated kinase 4 interacts with integrin alpha v beta 5 and regulates alpha v beta 5-mediated cell migration. J Cell Biol 158:1287–1297