doi:10.1038/nature11412

Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network*

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.

Breast cancer is one of the most common cancers with greater than 1,300,000 cases and 450,000 deaths each year worldwide. Clinically, this heterogeneous disease is categorized into three basic therapeutic groups. The oestrogen receptor (ER) positive group is the most numerous and diverse, with several genomic tests to assist in predicting outcomes for ER$^+$ patients receiving endocrine therapy[1,2]. The *HER2* (also called *ERBB2*) amplified group[3] is a great clinical success because of effective therapeutic targeting of *HER2*, which has led to intense efforts to characterize other DNA copy number aberrations[4,5]. Triple-negative breast cancers (TNBCs, lacking expression of ER, progesterone receptor (PR) and *HER2*), also known as basal-like breast cancers[6], are a group with only chemotherapy options, and have an increased incidence in patients with germline *BRCA1* mutations[7,8] or of African ancestry[9].

Most molecular studies of breast cancer have focused on just one or two high information content platforms, most frequently mRNA expression profiling or DNA copy number analysis, and more recently massively parallel sequencing[10–12]. Supervised clustering of mRNA expression data has reproducibly established that breast cancers encompass several distinct disease entities, often referred to as the intrinsic subtypes of breast cancer[13,14]. The recent development of additional high information content assays focused on abnormalities in DNA methylation, microRNA (miRNA) expression and protein expression, provide further opportunities to characterize more completely the molecular architecture of breast cancer. In this study, a diverse set of breast tumours were assayed using six different technology platforms. Individual platform and integrated pathway analyses identified many subtype-specific mutations and copy number changes that identify therapeutically tractable genomic aberrations and other events driving tumour biology.

Samples and clinical data

Tumour and germline DNA samples were obtained from 825 patients. Different subsets of patients were assayed on each platform: 466 tumours from 463 patients had data available on five platforms including Agilent mRNA expression microarrays ($n = 547$), Illumina Infinium DNA methylation chips ($n = 802$), Affymetrix 6.0 single nucleotide polymorphism (SNP) arrays ($n = 773$), miRNA sequencing ($n = 697$), and whole-exome sequencing ($n = 507$); in addition, 348 of the 466 samples also had reverse-phase protein array (RPPA) data ($n = 403$). Owing to the short median overall follow up (17 months) and the small number of overall survival events (93 out of 818), survival analyses will be presented in a later publication. Demographic and clinical characteristics are presented in Supplementary Table 1.

Significantly mutated genes in breast cancer

Overall, 510 tumours from 507 patients were subjected to whole-exome sequencing, identifying 30,626 somatic mutations comprised of 28,319 point mutations, 4 dinucleotide mutations, and 2,302 insertions/deletions (indels) (ranging from 1 to 53 nucleotides). The point mutations included 6,486 silent, 19,045 missense, 1,437 nonsense, 26 read-through, 506 splice-site mutations, and 819 mutations in RNA genes. Comparison to COSMIC and OMIM databases identified 619 mutations across 177 previously reported cancer genes. Of 19,045 missense mutations, 9,484 were predicted to have a high probability of being deleterious by Condel[15]. The MuSiC package[16], which determines the significance of the observed mutation rate of each gene based on the background mutation rate, identified 35 significantly mutated genes (excluding LOC or Ensembl gene IDs) by at least two tests (convolution and likelihood ratio tests) with false discovery rate (FDR) <5% (Supplementary Table 2).

In addition to identifying nearly all genes previously implicated in breast cancer (*PIK3CA*, *PTEN*, *AKT1*, *TP53*, *GATA3*, *CDH1*, *RB1*, *MLL3*, *MAP3K1* and *CDKN1B*), a number of novel significantly mutated genes were identified including *TBX3*, *RUNX1*, *CBFB*, *AFF2*, *PIK3R1*, *PTPN22*, *PTPRD*, *NF1*, *SF3B1* and *CCND3*. *TBX3*, which is mutated in ulnar-mammary syndrome and involved in mammary gland development[17], harboured 13 mutations (8 frame-shift indels,

*A list of participants and their affiliations appears at the end of the paper.

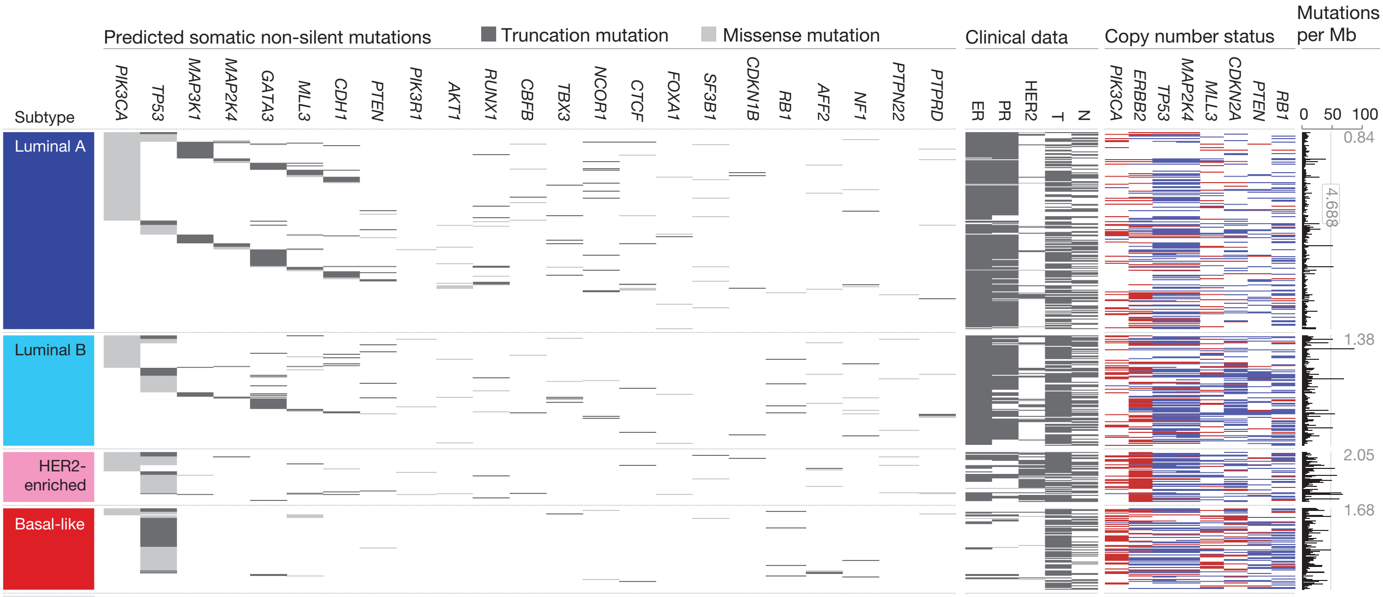
1 in-frame deletion, 1 nonsense, and 3 missense), suggesting a loss of function. Additionally, 2 mutations were found in *TBX4* and 1 mutation in *TBX5*, which are genes involved in Holt–Oram syndrome[18]. Two other transcription factors, *CTCF* and *FOXA1*, were at or near significance harbouring 13 and 8 mutations, respectively. *RUNX1* and *CBFB*, both rearranged in acute myeloid leukaemia and interfering with haematopoietic differentiation, harboured 19 and 9 mutations, respectively. *PIK3R1* contained 14 mutations, most of which clustered in the PIK3CA interaction domain similar to previously identified mutations in glioma[19] and endometrial cancer[20]. We also observed a statistically significant exclusion pattern among *PIK3R1*, *PIK3CA*, *PTEN* and *AKT1* mutations ($P = 0.025$). Mutation of splicing factor *SF3B1*, previously described in myelodysplastic syndromes[21] and chronic lymphocytic leukaemia[22], was significant with 15 non-silent mutations, of which 4 were a recurrent K700E substitution. Two protein tyrosine phosphatases (*PTPN22* and *PTPRD*) were also significantly mutated; frequent deletion/mutation of *PTPRD* is observed in lung adenocarcinoma[23].

Mutations and mRNA-expression subtype associations

We analysed the somatic mutation spectrum within the context of the four mRNA-expression subtypes, excluding the normal-like group owing to small numbers ($n = 8$) (Fig. 1). Several significantly mutated genes showed mRNA-subtype-specific (Supplementary Figs 1–3) and clinical-subtype-specific patterns of mutation (Supplementary Table 2). Significantly mutated genes were considerably more diverse and recurrent within luminal A and luminal B tumours than within basal-like and HER2-enriched (HER2E) subtypes; however, the overall mutation rate was lowest in luminal A subtype and highest in the basal-like and HER2E subtypes. The luminal A subtype harboured the most significantly mutated genes, with the most frequent being *PIK3CA* (45%), followed by *MAP3K1*, *GATA3*, *TP53*, *CDH1* and *MAP2K4*. Twelve per cent of luminal A tumours contained likely inactivating mutations in *MAP3K1* and *MAP2K4*, which represent two contiguous steps in the p38–JNK1 stress kinase pathway[24]. Luminal B cancers exhibited a diversity of significantly mutated genes, with *TP53* and *PIK3CA* (29% each) being the most frequent. The luminal tumour subtypes markedly contrasted with basal-like cancers where *TP53* mutations occurred in 80% of cases and the majority of the luminal significantly mutated gene repertoire, except *PIK3CA* (9%), were absent or near absent. The HER2E subtype, which has frequent *HER2* amplification (80%), had a hybrid pattern with a high frequency of *TP53* (72%) and *PIK3CA* (39%) mutations and a much lower frequency of other significantly mutated genes including *PIK3R1* (4%).

Intrinsic mRNA subtypes differed not only by mutation frequencies but also by mutation type. Most notably, *TP53* mutations in basal-like tumours were mostly nonsense and frame shift, whereas missense mutations predominated in luminal A and B tumours (Supplementary Fig. 1). Fifty-eight somatic *GATA3* mutations, some of which were previously described[25], were detected including a hotspot 2-base-pair deletion within intron 4 only in the luminal A subtype (13 out of 13 mutants) (Supplementary Fig. 2). In contrast, 7 out of 9 frame-shift mutations in exon 5 (DNA binding domain) occurred in luminal B cancers. *PIK3CA* mutation frequency and spectrum also varied by mRNA subtype (Supplementary Fig. 3); the recurrent *PIK3CA* E545K mutation was present almost exclusively within luminal A (25 out of 27) tumours. *CDH1* mutations were common (30 out of 36) within the lobular histological subtype and corresponded with lower *CDH1* mRNA (Supplementary Fig. 4) and protein expression. Finally, we identified 4 out of 8 somatic variants in *HER2* within lobular cancers, three of which were within the tyrosine kinase domain.

We performed analyses on a selected set of genes[26] using the normal tissue DNA data and detected a number of germline predisposing variants. These analyses identified 47 out of 507 patients with deleterious germline variants, representing nine different genes (*ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CHEK2*, *NBN*, *PTEN*, *RAD51C* and *TP53;* Supplementary Table 3), supporting the hypothesis that ~10% of sporadic breast cancers may have a strong germline contribution.



	PIK3CA	TP53	MAP3K1	MAP2K4	GATA3	MLL3	CDH1	PTEN	PIK3R1	AKT1	RUNX1	CBFB	TBX3	NCOR1	CTCF	FOXA1	SF3B1	CDKN1B	RB1	AFF2	NF1	PTPN22	PTPRD
All	36%	37%	8%	4%	11%	7%	7%	3%	3%	2%	4%	2%	3%	3%	3%	2%	2%	1%	2%	3%	3%	1%	2%
Luminal A	45%	12%	13%	7%	14%	8%	9%	4%	0.4%	4%	5%	2%	3%	5%	4%	2%	3%	1%	0.4%	1%	2%	0.4%	2%
Luminal B	29%	29%	5%	2%	15%	6%	5%	4%	2%	2%	2%	2%	4%	2%	2%	2%	0%	1%	3%	2%	4%	2%	4%
HER2-enriched	39%	72%	4%	2%	2%	7%	5%	2%	4%	2%	4%	2%	0%	0%	2%	2%	4%	2%	0%	5%	0%	5%	4%
Basal-like	9%	80%	0%	0%	2%	5%	0%	1%	0%	0%	0%	0%	1%	2%	1%	0%	1%	0%	4%	4%	2%	0%	1%

Percentages of cases with mutation by expression subtype

Figure 1 | Significantly mutated genes and correlations with genomic and clinical features. Tumour samples are grouped by mRNA subtype: luminal A ($n = 225$), luminal B ($n = 126$), HER2E ($n = 57$) and basal-like ($n = 93$). The left panel shows non-silent somatic mutation patterns and frequencies for significantly mutated genes. The middle panel shows clinical features: dark grey, positive or T2–4; white, negative or T1; light grey, N/A or equivocal. N, node status; T, tumour size. The right panel shows significantly mutated genes with frequent copy number amplifications (red) or deletions (blue). The far-right panel shows non-silent mutation rate per tumour (mutations per megabase, adjusted for coverage). The average mutation rate for each expression subtype is indicated. Hypermutated: mutation rates >3 s.d. above the mean (>4.688, indicated by grey line).

These data confirmed the association between the presence of germline *BRCA1* mutations and basal-like breast cancers[7,8].

Gene expression analyses (mRNA and miRNA)

Several approaches were used to look for structure in the mRNA expression data. We performed an unsupervised hierarchical clustering analysis of 525 tumours and 22 tumour-adjacent normal tissues using the top 3,662 variably expressed genes (Supplementary Fig. 5); SigClust analysis identified 12 classes (5 classes with >9 samples per class). We performed a semi-supervised hierarchical cluster analysis using a previously published 'intrinsic gene list'[14], which identified 13 classes (9 classes with >9 samples per class) (Supplementary Fig. 6). We also classified each sample using the 50-gene PAM50 model[14] (Supplementary Fig. 5). High concordance was observed between all three analyses; therefore, we used the PAM50-defined subtype predictor as a common classification metric. There were only eight normal-like and eight claudin-low tumours[27], thus we did not perform focussed analyses on these two subtypes.

MicroRNA expression levels were assayed via Illumina sequencing, using 1,222 miRBase[28] v16 mature and star strands as the reference database of miRNA transcripts/genes. Seven subtypes were identified by consensus non-negative matrix factorization (NMF) clustering using an abundance matrix containing the 25% most variable miRNAs (306 transcripts/genes or MIMATs (miRNA IDs)). These subtypes correlated with mRNA subtypes, ER, PR and HER2 clinical status (Supplementary Fig. 7). Of note, miRNA groups 4 and 5 showed high overlap with the basal-like mRNA subtype and contained many *TP53* mutations. The remaining miRNA groups (1–3, 6 and 7) were composed of a mixture of luminal A, luminal B and HER2E with little correlation with the PAM50 defined subtypes. With the exception of *TP53*—which showed a strong positive correlation—and *PIK3CA* and *GATA3*—which showed negative associations with groups 4 and 5, respectively—there was little correlation with mutation status and miRNA subtype.

DNA methylation

Illumina Infinium DNA methylation arrays were used to assay 802 breast tumours. Data from HumanMethylation27 (HM27) and HumanMethylation450 (HM450) arrays were combined and filtered to yield a common set of 574 probes used in an unsupervised clustering analysis, which identified five distinct DNA methylation groups (Supplementary Fig. 8). Group 3 showed a hypermethylated phenotype and was significantly enriched for luminal B mRNA subtype and under-represented for *PIK3CA*, *MAP3K1* and *MAP2K4* mutations. Group 5 showed the lowest levels of DNA methylation, overlapped with the basal-like mRNA subtype, and showed a high frequency of *TP53* mutations. HER2-positive ($HER2^+$) clinical status, or the HER2E mRNA subtype, had only a modest association with the methylation subtypes.

A supervised analysis of the DNA methylation and mRNA expression data was performed to compare DNA methylation group 3 ($N = 49$) versus all tumours in groups 1, 2 and 4 (excluding group 5, which consisted predominantly of basal-like tumours). This analysis identified 4,283 genes differentially methylated (3,735 higher in group 3 tumours) and 1,899 genes differentially expressed (1,232 downregulated); 490 genes were both methylated and showed lower expression in group 3 tumours (Supplementary Table 4). A DAVID (database for annotation, visualization and integrated discovery) functional annotation analysis identified 'extracellular region part' and 'Wnt signalling pathway' to be associated with this 490-gene set; the group 3 hypermethylated samples showed fewer *PIK3CA* and *MAP3K1* mutations, and lower expression of Wnt-pathway genes.

DNA copy number

A total of 773 breast tumours were assayed using Affymetrix 6.0 SNP arrays. Segmentation analysis and GISTIC were used to identify focal amplifications/deletions and arm-level gains and losses (Supplementary Table 5). These analyses confirmed all previously reported copy number variations and highlighted a number of significantly mutated genes including focal amplification of regions containing *PIK3CA*, *EGFR*, *FOXA1* and *HER2*, as well as focal deletions of regions containing *MLL3*, *PTEN*, *RB1* and *MAP2K4* (Supplementary Fig. 9); in all cases, multiple genes were included within each altered region. Importantly, many of these copy number changes correlated with mRNA subtype including characteristic loss of 5q and gain of 10p in basal-like cancers[5,29] and gain of 1q and/or 16q loss in luminal tumours[4]. NMF clustering of GISTIC segments identified five copy number clusters/groups that correlated with mRNA subtypes, ER, PR and HER2 clinical status, and *TP53* mutation status (Supplementary Fig. 10). In addition, this aCGH subtype classification was highly correlated with the aCGH subtypes recently defined by ref. 30 (Supplementary Fig. 11).

Reverse phase protein arrays

Quantified expression of 171 cancer-related proteins and phosphoproteins by RPPA was performed on 403 breast tumours[31]. Unsupervised hierarchical clustering analyses identified seven subtypes; one class contained too few cases for further analysis (Supplementary Fig. 12). These protein subtypes were highly concordant with the mRNA subtypes, particularly with basal-like and HER2E mRNA subtypes. Closer examination of the HER2-containing RPPA-defined subgroup showed coordinated overexpression of HER2 and EGFR with a strong concordance with phosphorylated HER2 (pY1248) and EGFR (pY992), probably from heterodimerization and cross-phosphorylation. Although there is a potential for modest cross reactivity of antibodies against these related total and phospho-proteins, the concordance of phosphorylation of HER2 and EGFR was confirmed using multiple independent antibodies.

In RPPA-defined luminal tumours, there was high protein expression of ER, PR, AR, BCL2, GATA3 and INPP4B, defining mostly luminal A cancers and a second more heterogeneous protein subgroup composed of both luminal A and luminal B cancers. Two potentially novel protein-defined subgroups were identified: reactive I consisted primarily of a subset of luminal A tumours, whereas reactive II consisted of a mixture of mRNA subtypes. These groups are termed 'reactive' because many of the characteristic proteins are probably produced by the microenvironment and/or cancer-activated fibroblasts including fibronectin, caveolin 1 and collagen VI. These two RPPA groups did not have a marked difference in the percentage tumour cell content when compared to each other, or the other protein subtypes, as assessed by SNP array analysis or pathological examination. In addition, supervised analyses of reactive I versus II groups using miRNA expression, DNA methylation, mutation, or DNA copy number data identified no significant differences between these groups, whereas similar supervised analyses using protein and mRNA expression identified many differences.

Multiplatform subtype discovery

To reveal higher-order structure in breast tumours based on multiple data types, significant clusters/subtypes from each of five platforms were analysed using a multiplatform data matrix subjected to unsupervised consensus clustering (Fig. 2). This 'cluster of clusters' (C-of-C) approach illustrated that basal-like cancers had the most distinct multiplatform signature as all the different platforms for the basal-like groups clustered together. To a great extent, the four major C-of-C subdivisions correlated well with the previously published mRNA subtypes (driven, in part, by the fact that the four intrinsic subtypes were one of the inputs). Therefore, we also performed C-of-C analysis with no mRNA data present (Supplementary Fig. 13) or with the 12 unsupervised mRNA subtypes (Supplementary Fig. 14), and in each case 4–6 groups were identified. Recent work identified ten copy-number-based subgroups in a 997 breast cancer set[30]. We evaluated

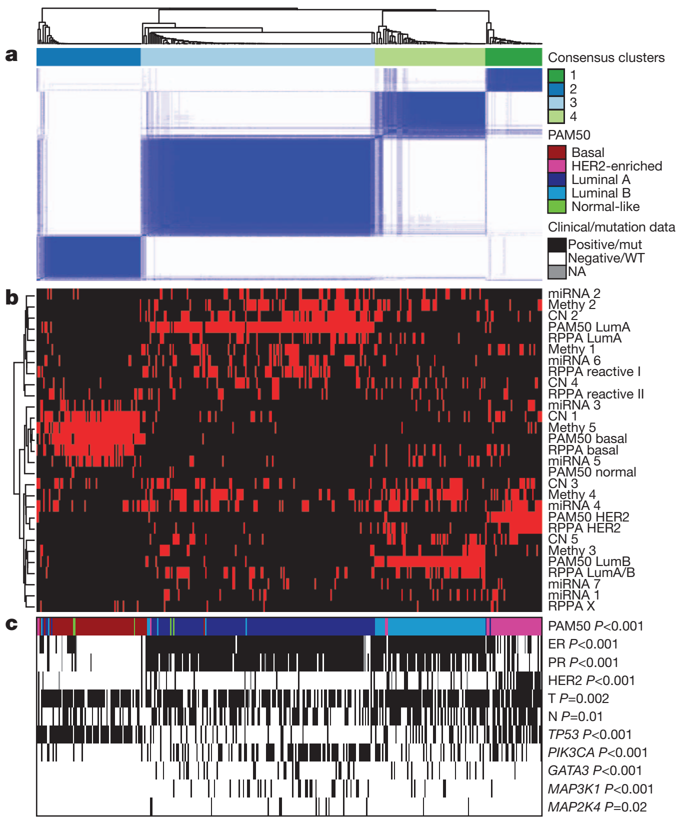


Figure 2 | Coordinated analysis of breast cancer subtypes defined from five different genomic/proteomic platforms. **a**, Consensus clustering analysis of the subtypes identifies four major groups (samples, $n = 348$). The blue and white heat map displays sample consensus. **b**, Heat-map display of the subtypes defined independently by miRNAs, DNA methylation, copy number (CN), PAM50 mRNA expression, and RPPA expression. The red bar indicates membership of a cluster type. **c**, Associations with molecular and clinical features. *P* values were calculated using a chi-squared test.

this classification in a C-of-C analysis instead of our five-class copy number subtypes, with either the PAM50 (Supplementary Fig. 15) or 12 unsupervised mRNA subtypes (Supplementary Fig. 16); each of these C-of-C classifications was highly correlated with PAM50 mRNA subtypes and with the other C-of-C analyses (Fig. 2). The transcriptional profiling and RPPA platforms demonstrated a high correlation with the consensus structure, indicating that the information content from copy number aberrations, miRNAs and methylation is captured at the level of gene expression and protein function.

Luminal/ER$^+$ summary analysis

Luminal/ER$^+$ breast cancers are the most heterogeneous in terms of gene expression (Supplementary Fig. 5), mutation spectrum (Fig. 1), copy number changes (Supplementary Fig. 9) and patient outcomes[1,14]. One of the most dominant features is high mRNA and protein expression of the luminal expression signature (Supplementary Fig. 5), which contains *ESR1*, *GATA3*, *FOXA1*, *XBP1* and *MYB*; the luminal/ER$^+$ cluster also contained the largest number of significantly mutated genes. Most notably, *GATA3* and *FOXA1* were mutated in a mutually exclusive fashion, whereas *ESR1* and *XBP1* were typically highly expressed but infrequently mutated. Mutations in *RUNX1* and its dimerization partner *CBFB* may also have a role in aberrant ER signalling in luminal tumours, as *RUNX1* functions as an ER 'DNA tethering factor'[32]. PARADIGM[33] analysis comparing luminal versus basal-like cancers further emphasized the presence of a hyperactivated FOXA1–ER complex as a critical network hub differentiating these two tumour subtypes (Supplementary Fig. 17).

A confirmatory finding here was the high mutation frequency of *PIK3CA* in luminal/ER$^+$ breast cancers[34,35]. Through multiple technology platforms, we examined possible relationships between *PIK3CA* mutation, *PTEN* loss, *INPP4B* loss and multiple gene and protein expression signatures of pathway activity. RPPA data demonstrated that pAKT, pS6 and p4EBP1, typical markers of phosphatidylinositol-3-OH kinase (PI(3)K) pathway activation, were not elevated in *PIK3CA*-mutated luminal A cancers; instead, they were highly expressed in basal-like and HER2E mRNA subtypes (the latter having frequent *PIK3CA* mutations) and correlated strongly with *INPP4B* and *PTEN* loss, and to a degree with *PIK3CA* amplification. Similarly, protein[36] and three mRNA signatures[37–39] of PI(3)K pathway activation were enriched in basal-like over luminal A cancers (Fig. 3a). This apparent disconnect between the presence of *PIK3CA* mutations and biomarkers of pathway activation has been previously noted[36].

Another striking luminal/ER$^+$ subtype finding was the frequent mutation of *MAP3K1* and *MAP2K4*, which represent two contiguous steps within the p38–JNK1 pathway[24,40]. These mutations are predicted to be inactivating, with *MAP2K4* also a target of focal DNA loss in luminal tumours (Supplementary Fig. 9). To explore the possible interplay between *PIK3CA*, *MAP3K* and *MAP2K4* signalling, MEMo analysis[41] was performed to identify mutually exclusive alterations targeting frequently altered genes likely to belong to the same pathway (Fig. 4). Across all breast cancers, MEMo identified a set of modules that highlight the differential activation events within the receptor tyrosine kinase (RTK)–PI(3)K pathway (Fig. 4a); mutations of *PIK3CA* were very common in luminal/ER$^+$ cancers whereas *PTEN* loss was more common in basal-like tumours. Almost all *MAP3K1*and *MAP2K4* mutations were in luminal tumours, yet *MAP3K1* and *MAP2K4* appeared almost mutually exclusive relative to one another.

The TP53 pathway was differentially inactivated in luminal/ER$^+$ breast cancers, with a low *TP53* mutation frequency in luminal A (12%) and a higher frequency in luminal B (29%) cancers (Fig. 1). In addition to *TP53* itself, a number of other pathway-inactivating events occurred including *ATM* loss and *MDM2* amplification (Figs 3b and 4b), both of which occurred more frequently within luminal B cancers. Gene expression analysis demonstrated that individual markers of functional TP53 (*GADD45A* and *CDKN1A*), and TP53 activity[42,43] signatures, were highest in luminal A cancers (Fig. 3b). These data indicate that the TP53 pathway remains largely intact in luminal A cancers but is often inactivated in the more aggressive luminal B cancers[44]. Other PARADIGM-based pathway differences driving luminal B versus luminal A included hyperactivation of transcriptional activity associated with *MYC* and *FOXM1* proliferation.

The critical retinoblastoma/RB1 pathway also showed mRNA-subtype-specific alterations (Fig. 3c). RB1 itself, by mRNA and protein expression, was detectable in most luminal cancers, with highest levels within luminal A. A common oncogenic event was cyclin D1 amplification and high expression, which preferentially occurred within luminal tumours, and more specifically within luminal B. In contrast, the presumed tumour suppressor *CDKN2C* (also called *p18*) was at its lowest levels in luminal A cancers, consistent with observations in mouse models[45]. Finally, RB1 activity signatures were also high in luminal cancers[46–48]. Luminal A tumours, which have the best prognosis, are the most likely to retain activity of the major tumour suppressors RB1 and TP53.

These genomic characterizations also provided clues for druggable targets. We compiled a drug target table in which we defined a target as a gene/protein for which there is an approved or investigational drug in human clinical trials targeting the molecule or canonical pathway (Supplementary Table 6). In luminal/ER$^+$ cancers, the high frequency of *PIK3CA* mutations suggests that inhibitors of this activated kinase or its signalling pathway may be beneficial. Other potential significantly mutated gene drug candidates include AKT1 inhibitors (11 out of 12 *AKT1* variants were luminal) and PARP inhibitors for *BRCA1*/*BRCA2* mutations. Although still unapproved as biomarkers, many potential copy-number-based drug targets

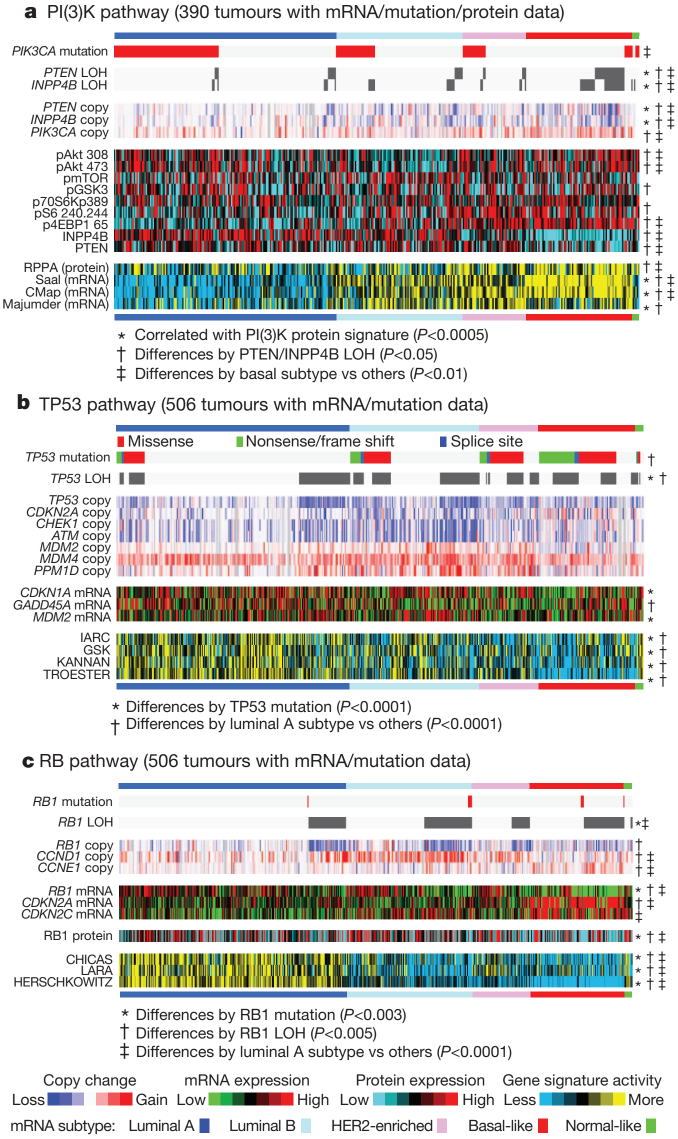


Figure 3 | Integrated analysis of the PI(3)K, TP53 and RB1 pathways. Breast cancer subtypes differ by genetic and genomic targeting events, with corresponding effects on pathway activity. **a**–**c**, For PI(3)K (**a**), TP53 (**b**) and RB1 (**c**) pathways, key genes were selected using prior biological knowledge. Multiple mRNA expression signatures for a given pathway were defined (details in Supplementary Methods; PI(3)K:Saal, PTEN loss in human breast tumours; CMap, PI(3)K/mTOR inhibitor treatment *in vitro*; Majumder, Akt overexpression in mouse model; TP53: IARC, expert-curated p53 targets; GSK, *TP53* mutant versus wild-type cell lines; KANNAN, TP53 overexpression *in vitro*; TROESTER, *TP53* knockdown *in vitro*; RB: CHICAS, *RB1* mouse knockout versus wild type; LARA, *RB1* knockdown *in vitro*; HERSCHKOWITZ, *RB1* loss of heterozygosity (LOH) in human breast tumours) and applied to the gene expression data, in order to score each tumour for relative signature activity (yellow, more active). The PI(3)K panel includes a protein-based (RPPA) proteomic signature. Tumours were ordered first by mRNA subtype, although specific ordering differs between the panels. *P* values were calculated by a Pearson's correlation or a Chi-squared test.

were identified including amplifications of fibroblast growth factor receptors (FGFRs) and *IGFR1*, as well as cyclin D1, *CDK4* and *CDK6*. A summary of the general findings in luminal tumours and the other subtypes is presented in Table 1.

HER2-based classifications and summary analysis

DNA amplification of *HER2* was readily evident in this study (Supplementary Fig. 9) together with overexpression of multiple HER2-amplicon-associated genes that in part define the HER2E mRNA subtype (Supplementary Fig. 5). However, not all clinically HER2$^+$ tumours are of the HER2E mRNA subtype, and not all tumours in the HER2E mRNA subtype are clinically HER2$^+$. Integrated analysis of the RPPA and mRNA data clearly identified a HER2$^+$ group (Supplementary Fig. 12). When the HER2$^+$ protein and HER2E mRNA subtypes overlapped, a strong signal of EGFR, pEGFR, HER2 and pHER2 was observed. However, only ~50% of clinically HER2$^+$ tumours fall into this HER2E-mRNA-subtype/HER2-protein group, the rest of the clinically HER2$^+$ tumours were observed predominantly in the luminal mRNA subtypes.

These data indicate that there exist at least two types of clinically defined HER2$^+$ tumours. To identify differences between these groups, a supervised gene expression analysis comparing 36 HER2E-mRNA-subtype/HER2$^+$ versus 31 luminal-mRNA-subtype/HER2$^+$ tumours was performed and identified 302 differentially expressed genes (q-value = 0%) (Supplementary Fig. 18 and Supplementary Table 7). These genes largely track with ER status but also indicated that HER2E-mRNA-subtype/HER2$^+$ tumours showed significantly higher expression of a number of RTKs including *FGFR4*, *EGFR*, *HER2* itself, as well as genes within the HER2 amplicon (including *GRB7*). Conversely, the luminal-mRNA-subtype/HER2$^+$ tumours showed higher expression of the luminal cluster of genes including *GATA3*, *BCL2* and *ESR1*. Further support for two types of clinically defined HER2$^+$ disease was evident in the somatic mutation data supervised by either mRNA subtype or ER status; *TP53* mutations were significantly enriched in HER2E or ER-negative tumours whereas *GATA3* mutations were only observed in luminal subtypes or ER$^+$ tumours.

Analysis of the RPPA data according to mRNA subtype identified 36 differentially expressed proteins (q-value <5%) (Supplementary Fig. 18G and Supplementary Table 8). The EGFR/pEGFR/HER2/pHER2 signal was again observed and present within the HER2E-mRNA-subtype/HER2$^+$ tumours, as was high pSRC and pS6; conversely, many protein markers of luminal cancers again distinguished the luminal-mRNA-subtype/HER2$^+$ tumours. Given the importance of clinical HER2 status, a more focused analysis was performed based on the RPPA-defined protein expression of HER2 (Supplementary Fig. 19)—the results strongly recapitulated findings from the RPPA and mRNA subtypes including a high correlation between HER2 clinical status, HER2 protein by RPPA, pHER2, EGFR and pEGFR. These multiple signatures, namely HER2E mRNA subtype, HER2 amplicon genes by mRNA expression, and RPPA EGFR/pEGFR/HER2/pHER2 signature, ultimately identify at least two groups/subtypes within clinically HER2$^+$ tumours (Table 1). These signatures represent breast cancer biomarker(s) that could potentially predict response to anti-HER2 targeted therapies.

Many therapeutic advances have been made for clinically HER2$^+$ disease. This study has identified additional somatic mutations that represent potential therapeutic targets within this group, including a high frequency of *PIK3CA* mutations (39%), a lower frequency of *PTEN* and *PIK3R1* mutations (Supplementary Table 6), and genomic losses of *PTEN* and *INPP4B*. Other possible druggable mutations included variants within HER family members including two somatic mutations in *HER2*, two within *EGFR*, and five within *HER3*. Pertuzumab, in combination with trastuzumab, targets the HER2–HER3 heterodimer[49]; however, these data suggest that targeting EGFR with HER2 could also be beneficial. Finally, the HER2E mRNA subtype typically showed high aneuploidy, the highest somatic mutation rate (Table 1), and DNA amplification of other potential therapeutic targets including FGFRs, *EGFR*, *CDK4* and cyclin D1.

Basal-like summary analysis

The basal-like subtype was discovered more than a decade ago by first-generation cDNA microarrays[13]. These tumours are often referred to as triple-negative breast cancers (TNBCs) because most basal-like tumours are typically negative for ER, PR and HER2.

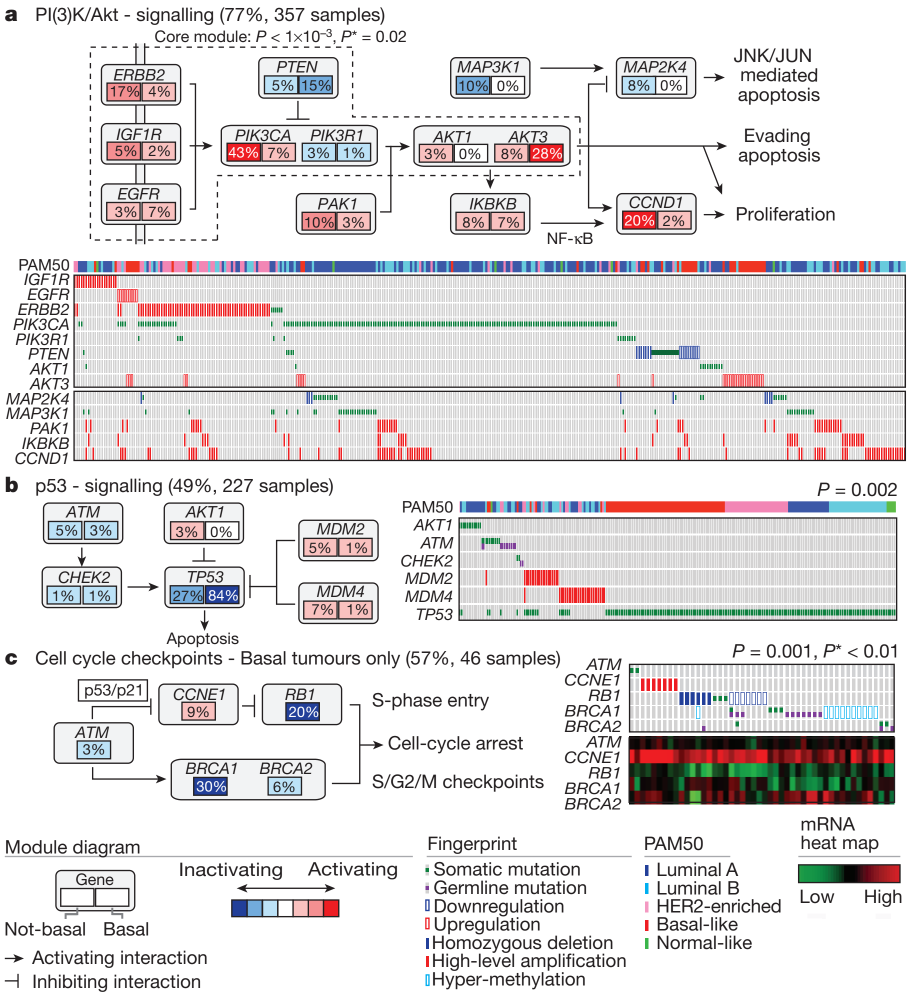


Figure 4 | Mutual exclusivity modules in cancer (MEMo) analysis. Mutual exclusivity modules are represented by their gene components and connected to reflect their activity in distinct pathways. For each gene, the frequency of alteration in basal-like (right box) and non-basal (left box) is reported. Next to each module is a fingerprint indicating what specific alteration is observed for each gene (row) in each sample (column). **a**, MEMo identified several overlapping modules that recapitulate the RTK–PI(3)K and p38–JNK1 signalling pathways and whose core was the top-scoring module. **b**, MEMo identified alterations to TP53 signalling as occurring within a statistically significant mutually exclusive trend. **c**, A basal-like only MEMo analysis identified one module that included ATM mutations, defects at *BRCA1* and *BRCA2*, and deregulation of the RB1 pathway. A gene expression heat map is below the fingerprint to show expression levels.

However, ~75% of TNBCs are basal-like with the other 25% comprised of all other mRNA subtypes[6]. In this data set, there was a high degree of overlap between these two distinctions with 76 TNBCs, 81 basal-like, and 65 that were both TNBCs and basal-like. Given the known heterogeneity of TNBCs, and that the basal-like subtype proved to be distinct on every platform, we chose to use the basal-like distinction for comparative analyses.

Basal-like tumours showed a high frequency of *TP53* mutations (80%)[9], which when combined with inferred TP53 pathway activity suggests that loss of TP53 function occurs within most, if not all, basal-like cancers (Fig. 3b). In addition to loss of *TP53*, a MEMo analysis reconfirmed that loss of *RB1* and *BRCA1* are basal-like features (Fig. 4c)[47,50]. *PIK3CA* was the next most commonly mutated gene (~9%); however, inferred PI(3)K pathway activity, whether from gene[37–39], protein[36], or high PI(3)K/AKT pathway activities, was highest in basal-like cancers (Fig. 3a). Alternative means of activating the PI(3)K pathway in basal-like cancers probably includes loss of *PTEN* and *INPP4B* and/or amplification of *PIK3CA*. A recent paper[12] performed exome sequencing of 102 TNBCs. Five of the top six most frequent TNBC mutations in ref. 12 were also observed at a similar frequency in our TNBC subset (*Myo3A* not present here); of those five, three passed our test as a significantly mutated gene in TNBCs (Supplementary Table 2).

Expression features of basal-like tumours include a characteristic signature containing keratins 5, 6 and 17 and high expression of genes associated with cell proliferation (Supplementary Fig. 5). A PARADIGM[33] analysis of basal-like versus luminal tumours emphasized the importance of hyperactivated FOXM1 as a transcriptional driver of this enhanced proliferation signature (Supplementary Fig. 17). PARADIGM also identified hyperactivated MYC and HIF1-α/ARNT network hubs as key regulatory features of basal-like cancers. Even though chromosome 8q24 is amplified across all subtypes (Supplementary Fig. 9), high MYC activation seems to be a basal-like characteristic[51].

Given the striking contrasts between basal-like and luminal/HER2E subtypes, we performed a MEMo analysis on basal-like tumours alone. The top-scoring module included *ATM* mutations, *BRCA1* and *BRCA2* inactivation, *RB1* loss and cyclin E1 amplification (Fig. 4c). Notably, these same modules were identified previously for serous ovarian cancers[41]. Furthermore, the basal-like (and TNBC) mutation spectrum was reminiscent of the spectrum seen in serous ovarian cancers[52] with only one gene (that is, *TP53*) at >10% mutation frequency. To explore possible similarities between serous ovarian and the breast basal-like cancers, we performed a number of analyses comparing ovarian versus breast luminal, ovarian versus breast basal-like, and breast basal-like versus breast luminal cancers

Table 1 | Highlights of genomic, clinical and proteomic features of subtypes

Subtype	Luminal A	Luminal B	Basal-like	HER2E
$ER^+/HER2^-$ (%)	87	82	10	20
$HER2^+$ (%)	7	15	2	68
TNBCs (%)	2	1	80	9
TP53 pathway	*TP53* mut (12%); gain of *MDM2* (14%)	*TP53* mut (32%); gain of *MDM2* (31%)	*TP53* mut (84%); gain of *MDM2* (14%)	*TP53* mut (75%); gain of *MDM2* (30%)
PIK3CA/PTEN pathway	*PIK3CA* mut (49%); *PTEN* mut/loss (13%); *INPP4B* loss (9%)	PIK3CA mut (32%) PTEN mut/loss (24%) INPP4B loss (16%)	PIK3CA mut (7%); PTEN mut/loss (35%); INPP4B loss (30%)	PIK3CA mut (42%); PTEN mut/loss (19%); INPP4B loss (30%)
RB1 pathway	Cyclin D1 amp (29%); *CDK4* gain (14%); low expression of *CDKN2C*; high expression of *RB1*	Cyclin D1 amp (58%); *CDK4* gain (25%)	*RB1* mut/loss (20%); cyclin E1 amp (9%); high expression of *CDKN2A*; low expression of *RB1*	Cyclin D1 amp (38%); *CDK4* gain (24%)
mRNA expression	High ER cluster; low proliferation	Lower ER cluster; high proliferation	Basal signature; high proliferation	HER2 amplicon signature; high proliferation
Copy number	Most diploid; many with quiet genomes; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (24%)	Most aneuploid; many with focal amp; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (51%); 8p11.23 amp (28%)	Most aneuploid; high genomic instability; 1q, 10p gain; 8p, 5q loss; *MYC* focal gain (40%)	Most aneuploid; high genomic instability; 1q, 8q gain; 8p loss; 17q12 focal *ERRB2* amp (71%)
DNA mutations	*PIK3CA* (49%); *TP53* (12%); *GATA3* (14%); *MAP3K1* (14%)	*TP53* (32%); *PIK3CA* (32%); *MAP3K1* (5%)	*TP53* (84%); *PIK3CA* (7%)	*TP53* (75%); *PIK3CA* (42%); *PIK3R1* (8%)
DNA methylation	–	Hypermethylated phenotype for subset	Hypomethylated	–
Protein expression	High oestrogen signalling; high MYB; RPPA reactive subtypes	Less oestrogen signalling; high FOXM1 and MYC; RPPA reactive subtypes	High expression of DNA repair proteins, PTEN and INPP4B loss signature (pAKT)	High protein and phospho-protein expression of EGFR and HER2

Percentages are based on 466 tumour overlap list. Amp, amplification; mut, mutation.

(Fig. 5). Comparing copy number landscapes, we observed several common features between ovarian and basal-like tumours including widespread genomic instability and common gains of 1q, 3q, 8q and 12p, and loss of 4q, 5q and 8p (Supplementary Fig. 20A). Using a more global copy number comparison, we examined the overall fraction of the genome altered and the overall copy number correlation of ovarian cancers versus each breast cancer mRNA subtype (Supplementary Fig. 20A, B); in both cases, basal-like tumours were the most similar to the serous ovarian carcinomas.

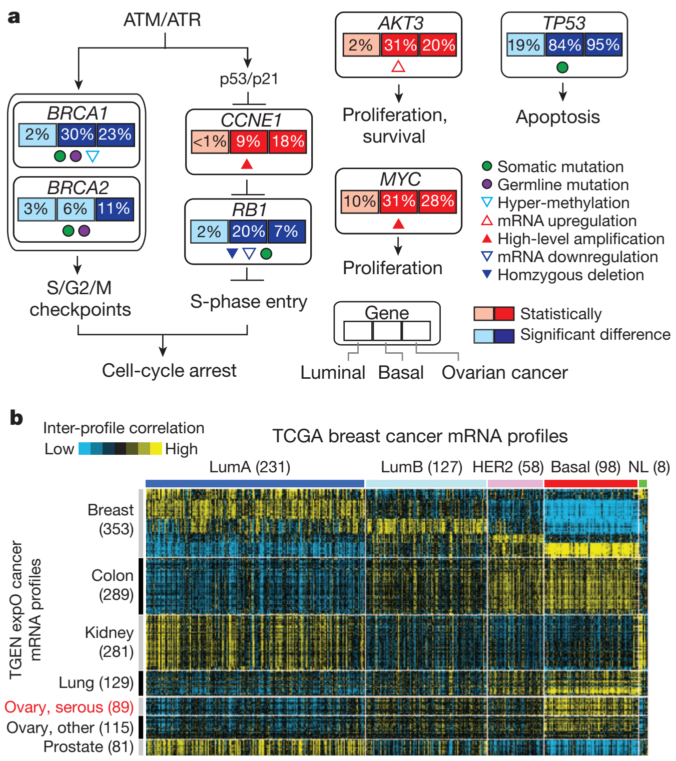


Figure 5 | Comparison of breast and serous ovarian carcinomas. **a**, Significantly enriched genomic alterations identified by comparing basal-like or serous ovarian tumours to luminal cancers. **b**, Inter-sample correlations (yellow, positive) between gene transcription profiles of breast tumours (columns; TCGA data, arranged by subtype) and profiles of cancers from various tissues of origin (rows; external 'TGEN expO' data set, GSE2109) including ovarian cancers.

We systematically looked for other common features between serous ovarian and basal-like tumours when each was compared to luminal. We identified: (1) *BRCA1* inactivation; (2) *RB1* loss and cyclin E1 amplification; (3) high expression of *AKT3*; (4) *MYC* amplification and high expression; and (5) a high frequency of *TP53* mutations (Fig. 5a). An additional supervised analysis of a large, external multitumour type transcriptomic data set (Gene Expression Omnibus accession GSE2109) was performed where each TCGA (The Cancer Genome Atlas) breast tumour expression profile was compared via a correlation analysis to that of each tumour in the multitumour set. Basal-like breast cancers clearly showed high mRNA expression correlations with serous ovarian cancers, as well as with lung squamous carcinomas (Fig. 5b). A PARADIGM analysis that calculates whether a gene or pathway feature is both differentially activated in basal-like versus luminal cancers and has higher overall activity across the TCGA ovarian samples was performed; this identified comparably high pathway activity of the HIF1-α/ARNT, MYC and FOXM1 regulatory hubs in both ovarian and basal-like cancers (Supplementary Fig. 20C). The common findings of *TP53*, *RB1* and *BRCA1* loss, with *MYC* amplification, strongly suggest that these are shared driving events for basal-like and serous ovarian carcinogenesis. This suggests that common therapeutic approaches should be considered, which is supported by the activity of platinum analogues and taxanes in breast basal-like and serous ovarian cancers.

Given that most basal-like cancers are TNBCs, finding new drug targets for this group is critical. Unfortunately, the somatic mutation repertoire for basal-like breast cancers has not provided a common target aside from *BRCA1* and *BRCA2*. Here we note that ~20% of basal-like tumours had a germline ($n = 12$) and/or somatic ($n = 8$) *BRCA1* or *BRCA2* variant, which suggests that one in five basal-like patients might benefit from PARP inhibitors and/or platinum compounds[53,54]. The copy number landscape of basal-like cancers showed multiple amplifications and deletions, some of which may provide therapeutic targets (Supplementary Table 6). Potential targets include losses of *PTEN* and *INPP4B*, both of which have been shown to sensitize cell lines to PI(3)K pathway inhibitors[55,56]. Interestingly, many of the components of the PI(3)K and RAS–RAF–MEK pathway were amplified (but not typically mutated) in basal-like cancers including *PIK3CA* (49%), *KRAS* (32%), *BRAF* (30%) and *EGFR* (23%). Other RTKs that are plausible drug targets and amplified in some basal-like cancers include *FGFR1*, *FGFR2*, *IGFR1*, *KIT*, *MET* and *PDGFRA*. Finally, the PARADIGM identification of high HIF1-α/ARNT pathway activity suggests that these malignancies might be susceptible to angiogenesis inhibitors and/or bioreductive drugs that become activated under hypoxic conditions.

Concluding remarks

The integrated molecular analyses of breast carcinomas that we report here significantly extends our knowledge base to produce a comprehensive catalogue of likely genomic drivers of the most common breast cancer subtypes (Table 1). Our novel observation that diverse genetic and epigenetic alterations converge phenotypically into four main breast cancer classes is not only consistent with convergent evolution of gene circuits, as seen across multiple organisms, but also with models of breast cancer clonal expansion and *in vivo* cell selection proposed to explain the phenotypic heterogeneity observed within defined breast cancer subtypes.

METHODS SUMMARY

Specimens were obtained from patients with appropriate consent from institutional review boards. Using a co-isolation protocol, DNA and RNA were purified. In total, 800 patients were assayed on at least one platform. Different numbers of patients were used for each platform using the largest number of patients available at the time of data freeze; 466 samples (463 patients) were in common across 5 out of 6 platforms (excluding RPPA) and 348 patients were in common on 6 out of 6 platforms. Technology platforms used include: (1) gene expression DNA microarrays[52]; (2) DNA methylation arrays; (3) miRNA sequencing; (4) Affymetrix SNP arrays; (5) exome sequencing; and (6) reverse phase protein arrays. Each platform, except for the exome sequencing, was used in a *de novo* subtype discovery analysis (Supplementary Methods) and then included in a single analysis to define an overall subtype architecture. Additional integrated across-platform computational analyses were preformed including PARADIGM[33] and MEMo[41].

Received 22 March; accepted 11 July 2012.

Published online 23 September; corrected online 3 October 2012.

1. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351,** 2817–2826 (2004).
2. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
3. Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235,** 177–182 (1987).
4. Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10,** 529–541 (2006).
5. Bergamaschi, A. *et al.* Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosom. Cancer* **45,** 1033–1040 (2006).
6. Perou, C. M. Molecular stratification of triple-negative breast cancers. *Oncologist* **16** (suppl. 1), 61–70 (2011).
7. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100,** 8418–8423 (2003).
8. Foulkes, W. D. *et al.* Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J. Natl Cancer Inst.* **95,** 1482–1485 (2003).
9. Carey, L. A. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J. Am. Med. Assoc.* **295,** 2492–2502 (2006).
10. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).
11. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
12. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–399 (2012).
13. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
14. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).
15. González-Pérez, A. & Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88,** 440–449 (2011).
16. Dees, N. D. *et al.* MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **22,** 1589–1598 (2012).
17. Bamshad, M. *et al.* Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16,** 311–315 (1997).
18. Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature Genet.* **15,** 21–29 (1997).
19. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
20. Cheung, L. W. *et al.* High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* **1,** 170–185 (2011).
21. Malcovati, L. *et al.* Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* **118,** 6239–6246 (2011).
22. Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365,** 2497–2506 (2011).
23. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455,** 1069–1075 (2008).
24. Johnson, G. L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **298,** 1911–1912 (2002).
25. Usary, J. *et al.* Mutation of GATA3 in human breast tumors. *Oncogene* **23,** 7669–7678 (2004).
26. Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **107,** 12629–12633 (2010).
27. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12,** R68 (2010).
28. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39,** D152–D157 (2011).
29. Weigman, V. J. *et al.* Basal-like breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res. Treat.* **133,** 865–880 (2011).
30. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346–352 (2012).
31. Hennessy, B. T. *et al.* A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clin. Proteomics* **6,** 129–151 (2010).
32. Daub, H. *et al.* Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol. Cell* **31,** 438–448 (2008).
33. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26,** 237–245 (2010).
34. Campbell, I. G. *et al.* Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res.* **64,** 7678–7681 (2004).
35. Bachman, K. E. *et al.* The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer Biol. Ther.* **3,** 772–775 (2004).
36. Stemke-Hale, K. *et al.* An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* **68,** 6084–6091 (2008).
37. Creighton, C. J. *et al.* Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER$^+$ breast cancer. *Breast Cancer Res.* **12,** R40 (2010).
38. Majumder, P. K. *et al.* mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nature Med.* **10,** 594–601 (2004).
39. Saal, L. H. *et al.* Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nature Genet.* **40,** 102–107 (2008).
40. Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nature Rev. Cancer* **9,** 537–549 (2009).
41. Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* **22,** 398–406 (2012).
42. Kannan, K. *et al.* DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene* **20,** 2225–2234 (2001).
43. Troester, M. A. *et al.* Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* **6,** 276 (2006).
44. Deisenroth, C., Thorner, A. R., Enomoto, T., Perou, C. M. & Zhang, Y. Mitochondrial Hep27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53. *Mol. Cell. Biol.* **30,** 3981–3993 (2010).
45. Pei, X. H. *et al.* CDK inhibitor p18^{INK4c} is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell* **15,** 389–401 (2009).
46. Chicas, A. *et al.* Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell* **17,** 376–387 (2010).
47. Herschkowitz, J. I., He, X., Fan, C. & Perou, C. M. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Res.* **10,** R75 (2008).
48. Lara, M. F. *et al.* Gene profiling approaches help to define the specific functions of retinoblastoma family in epidermis. *Mol. Carcinog.* **47,** 209–221 (2008).
49. Baselga, J. *et al.* Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N. Engl. J. Med.* **366,** 109–119 (2012).
50. Jiang, Z. *et al.* Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status. *J. Clin. Invest.* **120,** 3296–3309 (2010).
51. Chandriani, S. *et al.* A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS ONE* **4,** e6693 (2009).
52. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
53. Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet* **376,** 245–251 (2010).
54. Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361,** 123–134 (2009).

55. Fedele, C. G. *et al.* Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc. Natl Acad. Sci. USA* **107,** 22231–22236 (2010).
56. Gewinner, C. *et al.* Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* **16,** 115–125 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Sheth and S. Lucas for administrative coordination of TCGA activities, and C. Gunter for critical reading of the manuscript. This work was supported by the following grants from the USA National Institutes of Health: U24CA143883, U24CA143858, U24CA143840, U24CA143799, U24CA143835, U24CA143845, U24CA143882, U24CA143867, U24CA143866, U24CA143848, U24CA144025, U54HG003079, P50CA116201 and P50CA58223. Additional support was provided by the Susan G. Komen for the Cure, the US Department of Defense through the Henry M. Jackson Foundation for the Advancement of Military Medicine, and the Breast Cancer Research Foundation. The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of Defense, or US Government.

Author Contributions TCGA research network contributed collectively to this study. Biospecimens were provided by tissue source sites and processed by a Biospecimens Core Resource. Data generation and analyses were performed by genome sequencing centres, cancer genome characterization centres, and genome data analysis centres. RPPA analysis was performed at the MD Anderson Cancer Center in association with the genome data analysis centre. All data were released through the Data Coordinating Center. Project activities were coordinated by NCI and NHGRI project teams. We also acknowledge the following TCGA investigators of the Breast Analysis Working Group who contributed substantially to the analysis and writing of this manuscript: Project leaders, C.M.P., M.J.E.; manuscript coordinator, C.M.P., K.A.H.; data coordinator, K.A.H.; analysis coordinator, C.M.P., K.A.H.; DNA sequence analysis, D.C.K., L.D.; mRNA microarray analysis; K.A.H., C.F.; miRNA sequence analysis, A.G.R., A.C.; DNA methylation analysis, S. Malik, S. Mahurkar, P.W.L.; copy number analysis; A.D.C., M.M.; protein analysis, W.L., R.G.W.V., G.B.M.; pathway/integrated analysis, C.J.C., C.Y., J.M.S., C.C.B., G.C., C.S., S.R., I.S.; biospecimen core resource, T.L., J.B., J.M.G.; pathology and clinical expertise, T.A.K., H.H., R.J.M., J.N.I., T.S., F.W.

Author Information All of the primary sequence files are deposited in CGHub (https://cghub.ucsc.edu/); all other data including mutation annotation file are deposited at the Data Coordinating Center (http://cancergenome.nih.gov/). Sample lists, data matrices and supporting data can be found at http://tcga-data.nci.nih.gov/docs/publications/brca_2012/. The data can be explored via the ISB Regulome Explorer (http://explorer.cancerregulome.org/) and the cBio Cancer Genomics Portal (http://cbioportal.org). Data descriptions can be found at https://wiki.nci.nih.gov/display/TCGA/TCGA+Data+Primer and in Supplementary Methods. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of this paper is freely available to all readers. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.P. (cperou@med.unc.edu).

The Cancer Genome Atlas Network

Genome sequencing centres: Washington University in St Louis Daniel C. Koboldt[1], Robert S. Fulton[1], Michael D. McLellan[1], Heather Schmidt[1], Joelle Kalicki-Veizer[1], Joshua F. McMichael[1], Lucinda L. Fulton[1], David J. Dooling[1], Li Ding[1,2], Elaine R. Mardis[1,2,3], Richard K. Wilson[1,2,3]

Genome characterization centres: BC Cancer Agency Adrian Ally[4], Miruna Balasundaram[4], Yaron S. N. Butterfield[4], Rebecca Carlsen[4], Candace Carter[4], Andy Chu[4], Eric Chuah[4], Hye-Jung E. Chun[4], Robin J. N. Coope[4], Noreen Dhalla[4], Ranabir Guin[4], Carrie Hirst[4], Martin Hirst[4], Robert A. Holt[4], Darlene Lee[4], Haiyan I. Li[4], Michael Mayo[4], Richard A. Moore[4], Andrew J. Mungall[4], Erin Pleasance[4], A. Gordon Robertson[4], Jacqueline E. Schein[4], Arash Shafiei[4], Payal Sipahimalani[4], Jared R. Slobodan[4], Dominik Stoll[4], Angela Tam[4], Nina Thiessen[4], Richard J. Varhol[4], Natasja Wye[4], Thomas Zeng[4], Yongjun Zhao[4], Inanc Birol[4], Steven J. M. Jones[4], Marco A. Marra[4]; **Broad Institute** Andrew D. Cherniack[5], Gordon Saksena[5], Robert C. Onofrio[5], Nam H. Pho[5], Scott L. Carter[5], Steven E. Schumacher[5,6], Barbara Tabak[5,6], Bryan Hernandez[5], Jeff Gentry[5], Huy Nguyen[5], Andrew Crenshaw[5], Kristin Ardlie[5], Rameen Beroukhim[5,7,8], Wendy Winckler[5], Gad Getz[5], Stacey B. Gabriel[5], Matthew Meyerson[5,9,10]; **Brigham & Women's Hospital & Harvard Medical School** Lynda Chin[9,11], Peter J. Park[12], Raju Kucherlapati[13]; **University of North Carolina, Chapel Hill** Katherine A. Hoadley[14,15], J. Todd Auman[16,17], Cheng Fan[15], Yidi J. Turman[15], Yan Shi[15], Ling Li[15], Michael D. Topal[15,18], Xiaping He[14,15], Hann-Hsiang Chao[14,15], Aleix Prat[14,15], Grace O. Silva[14,15], Michael D. Iglesia[14,15], Wei Zhao[14,15], Jerry Usary[15], Jonathan S. Berg[14,15], Michael Adams[14], Jessica Booker[18], Junyuan Wu[15], Anisha Gulabani[15], Tom Bodenheimer[15], Alan P. Hoyle[15], Janae V. Simons[15], Matthew G. Soloway[15], Lisle E. Mose[15], Stuart R. Jefferys[15], Saianand Balu[15], Joel S. Parker[15], D. Neil Hayes[15,19], Charles M. Perou[14,15,18]; **University of Southern California/Johns Hopkins** Simeen Malik[20], Swapna Mahurkar[20], Hui Shen[20], Daniel J. Weisenberger[20], Timothy Triche Jr[20], Phillip H. Lai[20], Moiz S. Bootwalla[20], Dennis T. Maglinte[20], Benjamin P. Berman[20], David J. Van Den Berg[20], Stephen B. Baylin[21], Peter W. Laird[20]

Genome data analysis: Baylor College of Medicine Chad J. Creighton[22,23], Lawrence A. Donehower[22,23,24,25]; **Broad Institute** Gad Getz[26], Michael Noble[26], Doug Voet[26], Gordon Saksena[26], Nils Gehlenborg[12,26], Daniel DiCara[26], Juinhua Zhang[27], Hailei Zhang[26], Chang-Jiun Wu[28], Spring Yingchun Liu[26], Michael S. Lawrence[26], Lihua Zou[26], Andrey Sivachenko[26], Pei Lin[26], Petar Stojanov[26], Rui Jing[26], Juok Cho[26], Raktim Sinha[26], Richard W. Park[26], Marc-Danie Nazaire[26], Jim Robinson[26], Helga Thorvaldsdottir[26], Jill Mesirov[26], Peter J. Park[12,29,30], Lynda Chin[26,27]; **Institute for Systems Biology** Sheila Reynolds[31], Richard B. Kreisberg[31], Brady Bernard[31], Ryan Bressler[31], Timo Erkkila[32], Jake Lin[31], Vesteinn Thorsson[31], Wei Zhang[33], Ilya Shmulevich[31]; **Memorial Sloan-Kettering Cancer Center** Giovanni Ciriello[34], Nils Weinhold[34], Nikolaus Schultz[34], Jianjiong Gao[34], Ethan Cerami[34], Benjamin Gross[34], Anders Jacobsen[34], Rileen Sinha[34], B. Arman Aksoy[34], Yevgeniy Antipin[34], Boris Reva[34], Ronglai Shen[35], Barry S. Taylor[34], Marc Ladanyi[36], Chris Sander[34]; **Oregon Health & Science University** Pavana Anur[37], Paul T. Spellman[37]; **The University of Texas MD Anderson Cancer Center** Yiling Lu[38,39], Wenbin Liu[40], Roel R. G. Verhaak[40], Gordon B. Mills[38,39], Rehan Akbani[40], Nianxiang Zhang[40], Bradley M. Broom[40], Tod D. Casasent[40], Chris Wakefield[40], Anna K. Unruh[40], Keith Baggerly[40], Kevin Coombes[40], John N. Weinstein[40]; **University of California, Santa Cruz/Buck Institute** David Haussler[41,42], Christopher C. Benz[43], Joshua M. Stuart[41], Stephen C. Benz[41], Jingchun Zhu[41], Christopher C. Szeto[41], Gary K. Scott[43], Christina Yau[43], Evan O. Paull[41], Daniel Carlin[41], Christopher Wong[41], Artem Sokolov[41], Janita Thusberg[43], Sean Mooney[43], Sam Ng[41], Theodore C. Goldstein[41], Kyle Ellrott[41], Mia Grifford[41], Christopher Wilks[41], Singer Ma[41], Brian Craft[41]; **NCI** Chunhua Yan[44], Ying Hu[44], Daoud Meerzaman[44]

Biospecimen core resource: Nationwide Children's Hospital Biospecimen Core Resource Julie M. Gastier-Foster[45,46,47], Jay Bowen[47], Nilsa C. Ramirez[45,47], Aaron D. Black[47], Robert E. Pyatt[45,47], Peter White[46,47], Erik J. Zmuda[47], Jessica Frick[47], Tara M. Lichtenberg[47], Robin Brookens[47], Myra M. George[47], Mark A. Gerken[47], Hollie A. Harper[47], Kristen M. Leraas[47], Lisa J. Wise[47], Teresa R. Tabler[47], Cynthia McAllister[47], Thomas Barr[47], Melissa Hart-Kothari[47]

Tissue source sites: ABS-IUPUI Katie Tarvin[48], Charles Saller[49], George Sandusky[50], Colleen Mitchell[50]; **Christiana** Mary V. Iacocca[51], Jennifer Brown[51], Brenda Rabeno[51], Christine Czerwinski[51], Nicholas Petrelli[51]; **Cureline** Oleg Dolzhansky[52], Mikhail Abramov[53], Olga Voronina[54], Olga Potapova[54]; **Duke University Medical Center** Jeffrey R. Marks[55]; **The Greater Poland Cancer Centre** Wiktoria M. Suchorska[56], Dawid Murawa[56], Witold Kycler[56], Matthew Ibbs[56], Konstanty Korski[56], Arkadiusz Spychała[56], Paweł Murawa[56], Jacek J. Brzeziński[56], Hanna Perz[56], Radosław Łaźniak[56], Marek Teresiak[56], Honorata Tatka[56], Ewa Leporowska[56], Marta Bogusz-Czerniewicz[56,57], Julian Malicki[56,57], Andrzej Mackiewicz[56,57], Maciej Wiznerowicz[56,57]; **ILSBio** Xuan Van Le[58], Bernard Kohl[58], Nguyen Viet Tien[59], Richard Thorp[60], Nguyen Van Bang[61], Howard Sussman[62], Bui Duc Phu[61], Richard Hajek[63], Nguyen Phi Hung[64], Tran Viet The Phuong[65], Huynh Quyet Thang[66], Khurram Zaki Khan[60]; **International Genomics Consortium** Robert Penny[67], David Mallery[67], Erin Curley[67], Candace Shelton[67], Peggy Yena[67]; **Mayo Clinic** James N. Ingle[68], Fergus J. Couch[68], Wilma L. Lingle[68]; **MSKCC** Tari A. King[69]; **MD Anderson Cancer Center** Ana Maria Gonzalez-Angulo[38,70], Gordon B. Mills[70], Mary D. Dyer[70], Shuying Liu[70], Xiaolong Meng[70], Modesto Patangan[70]; **University of California San Francisco** Frederic Waldman[71,72], Hubert Stöppler[73]; **University of North Carolina** W. Kimryn Rathmell[15], Leigh Thorne[15,74], Mei Huang[15,74], Lori Boice[15,74], Ashley Hill[15]; **Roswell Park Cancer Institute** Carl Morrison[75], Carmelo Gaudioso[75], Wiam Bshara[75]; **University of Miami** Kelly Daily[76], Sophie C. Egea[76], Mark D. Pegram[76], Carmen Gomez-Fernandez[76]; **University of Pittsburgh** Rajiv Dhir[77], Rohit Bhargava[78], Adam Brufsky[78]; **Walter Reed National Military Medical Center** Craig D. Shriver[79], Jeffrey A. Hooke[79], Jamie Leigh Campbell[79], Richard J. Mural[80], Hai Hu[80], Stella Somiari[80], Caroline Larson[80], Brenda Deyarmin[80], Leonid Kvecher[80], Albert J. Kovatich[81]

Disease working group: Matthew J. Ellis[3,82,83], Tari A. King[69], Hai Hu[80], Fergus J. Couch[68], Richard J. Mural[80], Thomas Stricker[84], Kevin White[84], Olufunmilayo Olopade[85], James N. Ingle[68], Chunqing Luo[80], Yaqin Chen[80], Jeffrey R. Marks[55], Frederic Waldman[71,72], Maciej Wiznerowicz[56,57], Ron Bose[3,82,83], Li-Wei Chang[86], Andrew H. Beck[10], Ana Maria Gonzalez-Angulo[38,70]

Data coordination centre: Todd Pihl[87], Mark Jensen[87], Robert Sfeir[87], Ari Kahn[87], Anna Chu[87], Prachi Kothiyal[87], Zhining Wang[87], Eric Snyder[87], Joan Pontius[87], Brenda Ayala[87], Mark Backus[87], Jessica Walton[87], Julien Baboud[87], Dominique Berton[87], Matthew Nicholls[87], Deepak Srinivasan[87], Rohini Raman[87], Stanley Girshik[87], Peter Kigonya[87], Shelley Alonso[87], Rashmi Sanbhadti[87], Sean Barletta[87], David Pot[87]

Project team: National Cancer Institute Margi Sheth[88], John A. Demchok[88], Kenna R. Mills Shaw[88], Liming Yang[88], Greg Eley[89], Martin L. Ferguson[90], Roy W. Tarnuzzer[88], Jiashan Zhang[88], Laura A. L. Dillon[88], Kenneth Buetow[44], Peter Fielding[88]; **National Human Genome Research Institute** Bradley A. Ozenberger[91], Mark S. Guyer[91], Heidi J. Sofia[91], Jacqueline D. Palchik[91]

[1]The Genome Institute, Washington University, St Louis, Missouri 63108, USA. [2]Department of Genetics, Washington University, St Louis, Missouri 63110, USA. [3]Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA. [4]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z, Canada. [5]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [6]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [7]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. [8]Departments of Cancer Biology and Medical Oncology, and the Center for

Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [9]Department of Medical Oncology and the Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [10]Department of Pathology, Harvard Medical School, Boston, Massachusetts 02215, USA. [11]Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [12]The Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. [13]Department of Genetics, Harvard Medical School and Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [14]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [15]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [16]Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [17]Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [18]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [19]Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [20]USC Epigenome Center, University of Southern California, Los Angeles, California 90033, USA. [21]Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA. [22]Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. [23]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [24]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. [25]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA. [26]The Eli and Edythe L. Broad Institute of Massachusetts Institute Of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. [27]Institute for Applied Cancer Science, Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. [28]Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. [29]Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [30]Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA. [31]Institute for Systems Biology, Seattle, Washington 98109, USA. [32]Tampere University of Technology, Tampere, Finland. [33]Cancer Genomics Core Laboratory, MD Anderson Cancer Center, Houston, Texas 77030, USA. [34]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [35]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [36]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [37]Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA. [38]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [39]Kleberg Center for Molecular Markers, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [40]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [41]Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. [42]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. [43]Buck Institute for Research on Aging, Novato, California 94945, USA. [44]Center for Bioinformatics and Information Technology, National Cancer Institute, Rockville, Maryland 20852, USA. [45]The Ohio State University College of Medicine, Department of Pathology, Columbus, Ohio 43205, USA. [46]The Ohio State University College of Medicine, Department Pediatrics, Columbus, Ohio 43205, USA. [47]The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA. [48]ABS Inc. Indianapolis, Indiana 46204, USA. [49]ABS Inc. Wilmington, Delaware 19801, USA. [50]Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [51]Helen F. Graham Cancer Center, Christiana Care, Newark, Delaware 19713, USA. [52]Moscow City Clinical Oncology Dispensary 1 and the Central IHC Laboratory of the Moscow Health Department, Moscow 105005, Russia. [53]Russian Cancer Research Center, Moscow 115478, Russia. [54]Cureline, Inc., South San Francisco, California 94080, USA. [55]Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA. [56]The Greater Poland Cancer Centre, Poznań 61-866, Poland. [57]Poznan University of Medical Sciences, Poznań 61-701, Poland. [58]ILSBio, LLC, Chestertown, Maryland 21620, USA. [59]Ministry of Health, Hanoi, Vietnam. [60]ILSBio LLC, Karachi, Pakistan. [61]Hue Central Hospital, Hue City, Vietnam. [62]Stanford University Medical Center, Stanford, California 94305, USA. [63]Center for Minority Health Research, University of Texas, MD Anderson Cancer Center, Houston, Texas 07703, USA. [64]National Cancer Institute, Hanoi, Vietnam. [65]Ho Chi Minh City Cancer Center, Vietnam. [66]Can Tho Cancer Center, Can Tho, Vietnam. [67]International Genomics Consortium, Phoenix, Arizona 85004, USA. [68]Mayo Clinic, Rochester, Minnesota 55905, USA. [69]Department of Surgery, Breast Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [70]Department of Breast Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA. [71]University of California at San Francisco; San Francisco, California 94143, USA. [72]Cancer Diagnostics; Nichols Institute, Quest Diagnostics; San Juan Capistrano, California 92675, USA. [73]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94115, USA. [74]UNC Tissue Procurement Facility, Department of Pathology, UNC Lineberger Cancer Center, Chapel Hill, North Carolina 27599, USA. [75]Department of Pathology, Roswell Park Cancer Institute, Buffalo, New York 14263, USA. [76]Department of Pathology, University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. [77]University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. [78]Magee-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, USA. [79]Walter Reed National Military Medical Center, Bethesda, Maryland 20899-5600, USA. [80]Windber Research Institute, Windber, Pennsylvania 15963, USA. [81]MDR Global, LLC, Windber, Pennsylvania 15963, USA. [82]Breast Cancer Program, Washington University, St Louis, Missouri 63110, USA. [83]Department of Internal Medicine, Division of Oncology, Washington University, St Louis, Missouri 63110, USA. [84]Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. [85]Center for Clinical Cancer Genetics, The University of Chicago, Chicago, Illinois 60637, USA. [86]Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [87]SRA International, 4300 Fair Lakes Court, Fairfax, Virginia 22033, USA. [88]The Cancer Genome Atlas Program Office, Center for Cancer Genomics, National Cancer Institute, Bethesda, Maryland 20852, USA. [89]TCGA Consultant, Scimentis, LLC, Statham, Georgia 30666, USA. [90]MLF Consulting, Arlington, Massachusetts 02474, USA. [91]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.