



Published in final edited form as:

Nature. 2012 October 4; 490(7418): 61–70. doi:10.1038/nature11412.

Comprehensive molecular portraits of human breast tumors

The Cancer Genome Atlas Network

Summary

We analyzed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, mRNA 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 contributed by stromal/microenvironmental elements, and integrated analyses identified specific signaling pathways dominant in each molecular subtype including a HER2/p-HER2/HER1/p-HER1 signature within the HER2-Enriched expression subtype. Comparison of Basal-like breast tumors with high-grade Serous Ovarian tumors showed many molecular commonalities, suggesting a related etiology and similar therapeutic opportunities. The biologic 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 biologic subtypes of breast cancer.

Introduction

Breast cancer is one of the most common cancers with greater than 1,300,000 cases and 450,000 deaths each year world-wide. Clinically, this heterogeneous disease is categorized into three basic therapeutic groups. The Estrogen Receptor (ER) positive group are the most numerous and diverse, with several genomic tests to assist in predicting outcomes for ER+ patients receiving endocrine therapy^{1,2}. The *HER2/ERBB2* amplified group³ is a great clinical success because of effective therapeutic targeting of *HER2/ERBB2*, which has led to intense efforts to characterize other DNA copy number aberrations^{4,5}. Triple Negative Breast Cancers (TNBC), also known as Basal-like breast cancers⁶, are a group with only chemotherapy options, and an increased incidence in patients with germline *BRCA1* mutations^{7,8} or of African ancestry⁹.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Address Correspondence to: Charles M. Perou, PhD, Lineberger Comprehensive Cancer Center, 450 West Drive, CB# 7295, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, cperou@med.unc.edu.

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 expression and protein expression, provide further opportunities to more completely characterize the molecular architecture of breast cancer. In this study, a diverse set of breast tumors 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 tumor biology.

Samples and clinical data

Tumor and germline DNA samples were obtained from 825 patients. Different subsets of patients were assayed on each platform: 466 tumors 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 SNP arrays (n=773), microRNA 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). Due to the short median overall follow up (17 months) and the small number of overall survival events (93/818), survival analyses will be presented in a later publication. Demographic and clinical characteristics are presented in Supplemental Table 1.

Significantly mutated genes in breast cancer

Overall, 510 tumors 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 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 Condel¹⁵. The MuSiC package (Dees *et al.*, submitted), which determines the significance of the observed mutation rate of each gene based on the background mutation rate, identified 35 Significantly Mutated Genes/SMG (excluding LOC or ENSG genes) by at least two tests (Convolution and Likelihood Ratio tests) with FDR < 5% (Supplemental Table 2).

In addition to identifying nearly all genes previously implicated in breast cancer (*PIK3CA*, *PTEN*, *AKT1*, *TP53*, *GATA3*, *CDH1*, *RBI*, *MLL3*, *MAP3K1* and *CDKN1B*), a number of novel SMGs 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 development¹⁶, harbored 13 mutations (8 frame-shift indels, 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 Syndrome¹⁷. Two other transcription factors, *CTCF* and *FOXA1*, were at or near significance harboring 13 and 8 mutations, respectively. *RUNX1* and *CBFB*, both rearranged in Acute Myeloid Leukemia and interfering with hematopoietic differentiation, harbored 19 and 9 mutations, respectively. *PIK3RI* contained 14 mutations, most of which clustered in the PIK3CA interaction domain similar to previously identified mutations in glioma¹⁸ and endometrial cancer¹⁹. We also observed a statistically significant exclusion pattern among *PIK3RI*, *PIK3CA*, *PTEN*, and *AKT1* mutations ($P = 0.025$). Mutation of splicing factor *SF3B1*, previously described in Myelodysplastic Syndromes²⁰ and Chronic Lymphocytic Leukemia²¹, 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 adenocarcinoma²².

Association between mutations and mRNA-expression subtypes

We analyzed the somatic mutation spectrum within the context of the four mRNA-expression subtypes, excluding the Normal-like group due to small numbers ($n=8$) (Figure 1). Several SMGs showed mRNA-subtype (Supplemental Figures 1–3) and clinical-subtype specific patterns of mutation (Supplemental Table 2). SMGs were considerably more diverse and recurrent within Luminal A and Luminal B tumors than within Basal-like and 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 harbored the most SMGs, with the most frequent being *PIK3CA* (45%), followed by *MAP3K1*, *GATA3*, *TP53*, *CDH1*, and *MAP2K4*. 12% of Luminal A tumors contained likely inactivating mutations in *MAP3K1* and *MAP2K4*, which represent two contiguous steps in the p38/JNK1 stress kinase pathway²³. Luminal B cancers exhibited a diversity of SMGs with *TP53* and *PIK3CA* (29% each) being the most frequent. The luminal tumor subtypes markedly contrasted with Basal-like cancers where *TP53* mutations occurred in 80% of cases and the majority of the Luminal SMG repertoire, except *PIK3CA* (9%), were absent or near absent. The HER2-Enriched subtype (HER2E), which has frequent *HER2/ERBB2* amplification (80%), had a hybrid pattern with a high frequency of *TP53* (72%) and *PIK3CA* (39%) mutations and a much lower frequency of other SMGs including *PIK3RI* (4%).

Intrinsic mRNA-subtypes differed not only by mutation frequencies, but also by mutation type. Most notably, *TP53* mutations in Basal-like tumors were mostly nonsense and frame-shift, while missense mutations predominated in Luminal A and B tumors (Supplemental Figure 1). 58 somatic *GATA3* mutations, some of which were previously described²⁴, were detected including a hot spot 2bp deletion within intron 4 only in the Luminal A subtype (13/13 mutants) (Supplemental Figure 2). In contrast, 7/9 frame-shift mutations in exon 5 (DNA binding domain) occurred in Luminal B cancers. *PIK3CA* mutation frequency and spectrum also varied by mRNA-subtype (Supplemental Figure 3); the recurrent *PIK3CA* E545K mutation was present almost exclusively within Luminal A (25/27) tumors. *CDH1* mutations were common (30/36) within the lobular histologic subtype and corresponded with lower *CDH1* mRNA (Supplemental Figure 4) and protein expression. Finally, we identified 4/8 somatic variants in *HER2/ERBB2* within lobular cancers, 3 of which were within the tyrosine kinase domain.

We performed analyses on a selected set of genes²⁵ using the normal tissue DNA data and detected a number of germline predisposing variants. These analyses identified 47/507 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. These data confirmed the association between the presence of germline *BRCA1* mutations and Basal-like breast cancers^{7,8}.

Gene expression analyses (mRNA and microRNA)

Several approaches were used to look for structure in the mRNA expression data. We performed an unsupervised hierarchical clustering analysis of 525 tumors and 22 tumor-adjacent normal tissues using the top 3,662 variably expressed genes (Supplemental Figure 5); SigClust analysis identified 12 classes (5 classes with >9 samples/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/class) (Supplemental Figure 6). We also classified each sample using the 50-gene PAM50 model¹⁴ (Supplemental Figure 5). High concordance was observed between all three, therefore, we used the PAM50-defined subtype predictor as a common classification metric. There were only 8 Normal-like and 8 Claudin-low tumors²⁶, thus no analyses focused on these two subtypes were performed.

MicroRNA expression levels were assayed via Illumina sequencing, using 1222 miRBase²⁷ v16 mature and star strands as the reference database of microRNA transcripts/genes. Seven subtypes were identified by consensus NMF clustering using an abundance matrix containing the 25% most variable microRNAs (306 transcripts/genes or MIMATs). These subtypes correlated with mRNA-subtypes, ER, PR and HER2 clinical status (Supplemental Figure 7). Of note, microRNA groups 4 and 5 showed high overlap with the Basal-like mRNA subtype and contained many *TP53* mutations. The remaining microRNA groups (1–3, 6 and 7) were composed of a mixture of Luminal A, Luminal B and HER2-positive tumors with little correlation with the PAM50 defined subtypes. With the exception of *TP53*, which showed a strong positive correlation, and *PIK3CA* and *GATA3* that showed negative associations with groups 4 and 5 respectively, there was little correlation with mutation status and microRNA subtype.

DNA methylation

Illumina Infinium DNA methylation arrays were used to assay 802 breast tumors. 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 (Supplemental Figure 8). Group 3 showed a hyper-methylated phenotype and was significantly enriched for Luminal B mRNA-subtype and under-represented for *PIK3CA* and *MAP3K1/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 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 the DNA methylation Group 3 (N=49) versus all tumors in groups 1, 2, and 4 (excluding Group 5 that consisted predominantly of Basal-like tumors). This analysis identified 4,283 genes differentially methylated (3,735 higher in Group 3 tumors) and 1,899 genes differentially expressed (1,232 downregulated); 490 genes were both methylated and showed lower expression in Group 3 tumors (Supplemental Table 4). A DAVID functional annotation analysis identified 'Extracellular region part' and 'Wnt signaling pathway' to be associated with this 490 gene-set; the Group 3 hyper-methylated samples showed fewer *PIK3CA* and *MAP3K1* mutations, and lower expression of Wnt-pathway genes.

DNA copy number

773 breast tumors 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 (Supplemental Table 5). These analyses confirmed all previously reported copy number variations and highlighted a number of SMGs including focal amplification of regions containing *PIK3CA*, *HER1/EGFR*, *FOXA1*, and *HER2/ERBB2*, as well as focal deletions of regions containing *MLL3*, *PTEN*, *RB1* and *MAP2K4* (Supplemental Figure 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^{5,28} and gain of 1q and/or 16q loss in Luminal tumors⁴. 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 (Supplemental Figure 10). In addition, this aCGH subtype classification was highly correlated with the aCGH subtypes recently defined by Curtis et al.²⁹(Supplemental Figure 11).

Reverse Phase Protein Arrays (RPPA)

Quantified expression of 171 cancer-related proteins and phospho-proteins by RPPA was performed on 403 breast tumors³⁰. Unsupervised hierarchical clustering analyses identified seven subtypes; one class contained too few cases for further analysis (Supplemental Figure 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 HER1 with a strong concordance with phosphorylated HER2 (pY1248) and HER1 (pY992), likely from heterodimerization and cross-phosphorylation. While there is a potential for modest cross reactivity of antibodies against these related total and phosphoproteins, the concordance of phosphorylation of HER2 and HER1 was confirmed using multiple independent antibodies.

In RPPA-defined Luminal tumors, 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 tumors, whereas Reactive II consisted of a mixture of mRNA-subtypes. These groups are termed ‘Reactive’ because many of the characteristic proteins are likely 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 % tumor cell content versus each other, or the other protein subtypes, as assessed by SNP array analysis or pathological examination. In addition, supervised analyses of Reactive I vs. II groups using microRNA expression, DNA methylation, mutation, or DNA copy number data identified no significant differences between these groups, while similar supervised analyses using protein and mRNA expression identified many differences.

Multi-platform subtype discovery

To reveal higher order structure in breast tumors based on multiple data types, significant clusters/subtypes from each of five platforms were analyzed using a multi-platform data matrix subjected to unsupervised Consensus Clustering (Figure 2). This “Cluster of Clusters” (C of C) approach illustrated that Basal-like cancers had the most distinct multi-platform signature since 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 (Supplemental Figure 13) or with the 12 unsupervised mRNA subtypes (Supplemental Figure 14), and in each case 4–6 groups were identified. Recent work by Curtis et al. identified 10 copy number based subgroups in a 997 breast cancer set²⁹. We evaluated this classification in a C of C analysis instead of our 5-class copy number subtypes, with either the PAM50 (Supplemental Figure 15) or 12 unsupervised mRNA subtypes (Supplemental Figure 16); each of these C of C classifications were highly correlated with PAM50 mRNA-subtypes and with the other C of C analyses (Figure 2). The transcriptional profiling and RPPA platforms demonstrated a high correlation with the consensus structure suggesting that the information content from copy number aberrations, microRNAs, and methylation is captured at the level of gene expression and protein function.

Luminal/ER-positive summary analysis

Luminal/ER-positive breast cancers are the most heterogeneous in terms of gene expression (Supplemental Figure 5), mutation spectrum (Figure 1), copy number changes (Supplemental Figure 9) and patient outcomes^{1,14}. One of the most dominant features is high mRNA and protein expression of the luminal expression signature (Supplemental Figure 5), which contains *ESR1*, *GATA3*, *FOXA1*, *XBPI* and *cMYB*; the Luminal/ER-positive cluster also contained the largest number of SMGs. Most notably, *GATA3* and *FOXA1* were mutated in a mutually exclusive fashion, while *ESR1* and *XBPI* were typically highly expressed but infrequently mutated. Mutations in *RUNX1* and its dimerization partner *CBFB* may also play a role in aberrant ER-signaling in Luminals as *RUNX1* functions as an ER “DNA tethering factor”³¹. PARADIGM³² analysis comparing Luminal vs. Basal-like cancers further emphasized the presence of a hyperactivated FOXA1/ER complex as a critical network hub differentiating these two tumor subtypes (Supplemental Figure 17).

A confirmatory finding here was the high mutation frequency of *PIK3CA* in Luminal/ER-positive breast cancers^{33,34}. 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 PI3K 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, protein³⁵ and three mRNA signatures^{36–38} of PI3K pathway activation were enriched in Basal-like over Luminal A cancers (Figure 3A). This apparent disconnect between the presence of *PIK3CA* mutations and biomarkers of pathway activation has been previously noted³⁵.

Another striking Luminal/ER-positive subtype finding was the frequent mutation of *MAP3K1* and *MAP2K4*, which represent two contiguous steps within the p38/JNK1 pathway^{23,39}. These mutations are predicted to be inactivating, with *MAP2K4* also a target of focal DNA loss in Luminal tumors (Supplemental Figure 9). To explore the possible interplay between *PIK3CA* and *MAP3K1/MAP2K4* signaling, MEMo analysis⁴⁰ was performed to identify mutually exclusive alterations targeting frequently altered genes likely to belong to the same pathway (Figure 4). Across all breast cancers, MEMo identified a set of modules that highlight the differential activation events within the Receptor Tyrosine Kinase (RTK)/PI3K pathway (Figure 4A); mutations of *PIK3CA* were very common in Luminal/ER-positive cancers while *PTEN* loss was more common in Basal-like tumors. Almost all *MAP3K1/MAP2K4* mutations were in Luminal tumors, yet *MAP3K1* and *MAP2K4* appeared almost mutually exclusive relative to one another.

The TP53 pathway was differentially inactivated in Luminal/ER-positive breast cancers, with a low *TP53* mutation frequency in Luminal A (12%) and a higher frequency in Luminal B (29%) (Figure 1). In addition to *TP53* itself, a number of other pathway-inactivating events occurred including *ATM* loss and *MDM2* amplification (Figure 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^{41,42} signatures, were highest in Luminal A (Figure 3B). These data suggest that the TP53 pathway remains largely intact in Luminal A but is often inactivated in the more aggressive Luminal B⁴³. Other PARADIGM-based pathway differences driving Luminal B versus Luminal A included hyperactivation of transcriptional activity associated with *cMYC* and *FOXM1*/proliferation.

The critical retinoblastoma/RB1 pathway also showed mRNA-subtype specific alterations (Figure 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 tumors, and more specifically within Luminal B. In contrast, the presumed tumor suppressor p18/*CDKN2C* was at its lowest levels in Luminal A, consistent with observations from mouse models⁴⁴. Lastly, RB1 activity signatures were also high in Luminal cancers^{45–47}. Luminal A tumors, which have the best prognosis, are the most likely to retain activity of the major tumor 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 (Supplemental Table 6). In Luminal/ER-positive cancers, the high frequency of *PIK3CA* mutations suggests that inhibitors of this activated kinase or its signaling pathway may be beneficial. Other potential SMG drug candidates include AKT1 inhibitors (11/12 AKT1 variants were Luminal) and PARP inhibitors for BRCA1/2 mutations. Though still unapproved as biomarkers, many potential copy number-based drug targets were identified including amplifications of FGFRs and IGFR1, as well as Cyclin D1/CDK4/CDK6. A summary of the general findings in Luminal tumors and the other subtypes is presented in Table 1.

HER2-based classifications and summary analysis

DNA amplification of *HER2/ERBB2* was readily evident in this study (Supplemental Figure 9), together with overexpression of multiple HER2-amplicon associated genes that in part, define the HER2-Enriched (HER2E) mRNA-subtype (Supplemental Figure 5). However, not all clinically HER2+ tumors are of the HER2E mRNA-subtype, and not all tumors in the HER2E mRNA-subtype are clinically HER2+. Integrated analysis of the RPPA and mRNA data clearly identified a HER2+ group (Supplemental Figure 12). When the HER2-positive protein and HER2E mRNA-subtypes overlapped, a strong signal of HER1, p-HER1, HER2, and p-HER2 was observed. However, only ~50% of clinically HER2-positive tumors fall into this HER2E-mRNA-subtype/HER2-protein group, the rest of the clinically HER2-positive tumors were observed predominantly in the Luminal mRNA subtypes.

These data suggest that there exist at least two types of clinically defined HER2-positive tumors. To identify differences between these groups, a supervised gene expression analysis comparing 36 HER2E-mRNA-subtype/HER2-positive versus 31 Luminal-mRNA-subtype/HER2-positive tumors was performed and identified 302 differentially expressed genes (q-value = 0%) (Supplemental Figure 18, Supplemental Table 7). These genes largely track with ER status but also indicated that HER2E-mRNA-subtype/HER2-positive tumors showed significantly higher expression of a number of RTKs including *FGFR4*, *HER1/EGFR*, *HER2* itself, as well as genes within the HER2-amplicon (i.e. *GRB7*). Conversely, the Luminal-mRNA-subtype/HER2-positive tumors 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 tumors while *GATA3* mutations were only observed in Luminal subtypes or ER-positive tumors.

Analysis of the RPPA data according to mRNA-subtype identified 36 differentially expressed proteins (q-value <5%) (Supplemental Figure 18G, Supplemental Table 8). The HER1/p-HER1/HER2/p-HER2 signal was again observed and present within the HER2E-mRNA-subtype/HER2-positive tumors, as was high p-SRC and p-S6; conversely, many protein markers of Luminal cancers again distinguished the Luminal-mRNA-subtype/HER2-positive tumors. Given the importance of clinical HER2 status, a more focused analysis was

performed based upon the RPPA-defined protein expression of HER2 (Supplemental Figure 19); the results strongly recapitulated findings from the RPPA and mRNA-subtypes including a high correlation between HER2 clinical status, HER2 protein by RPPA, p-HER2, HER1, and p-HER1. These multiple signatures, namely HER2E-mRNA-subtype, HER2-amplicon genes by mRNA expression, and RPPA HER1/p-HER1/HER2/p-HER2 signature, ultimately identify at least two groups/subtypes within clinically HER2+ tumors (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-positive 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 (Supplemental 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 *HER1/EGFR*, and five within *HER3*. Pertuzumab, in combination with trastuzumab, targets the HER2-HER3 heterodimer⁴⁸, however, these data suggest that targeting HER1 with HER2 could also be beneficial. Lastly, 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*, *HER1/EGFR*, *CDK4* and *Cyclin D1*.

Basal-like summary analysis

The Basal-like subtype was discovered more than a decade ago by first-generation cDNA microarrays¹³. These tumors are often referred to as Triple-Negative Breast Cancers (TNBC) because most Basal-like tumors are typically negative for ER, PR and HER2. However, ~75% of TNBC are Basal-like with the other 25% comprised of all other mRNA-subtypes⁶. In this data set, there was a high degree of overlap between these two distinctions with 76 TNBC, 81 Basal-like, and 65 that were both TNBC and Basal-like. Given the known heterogeneity of TNBC, and that the Basal-like subtype proved to be distinct on every platform, we choose to use the Basal-like distinction for comparative analyses.

Basal-like tumors 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 (Figure 3B). In addition to loss of *TP53*, a MEMO analysis reconfirmed that loss of *RBI* and *BRCA1* are Basal-like features (Figure 4C)^{46,49}. *PIK3CA* was the next most commonly mutated gene (~9%); however, inferred PI3K pathway activity, whether from gene³⁶⁻³⁸, protein³⁵, or high PI3K/AKT pathway activities, was highest in Basal-like cancers (Figure 3A). Alternative means of activating the PI3K-pathway in Basal-like cancers likely includes loss of *PTEN* and *INPP4B* and/or amplification of *PIK3CA*. A recent paper from Shah et al.¹² performed exome sequencing of 102 TNBC. Five of the top six most frequent TNBCs mutations in Shah et al. 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 (Supplemental Table 2).

Expression features of Basal-like tumors include a characteristic signature containing Keratins 5, 6 and 17 and high expression of genes associated with cell proliferation (Supplemental Figure 5). A PARADIGM³² analysis of Basal-like versus Luminal tumors emphasized the importance of hyperactivated FOXM1 as a transcriptional driver of this enhanced proliferation signature (Supplemental Figure 17). PARADIGM also identified hyperactivated cMYC and HIF1 α /ARNT network hubs as key regulatory features of Basal-like cancers. Even though chromosome 8q24 is amplified across all subtypes (Supplemental Figure 9), high cMYC activation appears to be a Basal-like characteristic⁵⁰.

Given the striking contrasts between Basal-like and Luminal/HER2E subtypes, we performed a MEMo analysis on Basal-like tumors alone. The top-scoring module included *ATM* mutations, *BRCA1* and *BRCA2* inactivation, *RBI* loss and *Cyclin E1* amplification (Figure 4C). Interestingly, these same modules were identified previously for Serous Ovarian cancers⁴⁰. Furthermore, the Basal-like (and TNBC) mutation spectrum was reminiscent of the spectrum seen in Serous Ovarian cancers⁵¹ with only one gene (i.e. *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 vs. breast Luminal, Ovarian vs. breast Basal-like, and breast Basal-like vs. breast Luminal (Figure 5). Comparing copy number landscapes, we observed several common features between Ovarian and Basal-like tumors including widespread genomic instability and common gains of 1q, 3q, 8q and 12p, and loss of 4q, 5q and 8p (Supplemental Figure 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 (Supplemental Figure 20A, B); in both cases, Basal-like tumors were the most similar to the Serous Ovarian carcinomas.

We systematically looked for other common features between Serous Ovarian and Basal-like when each was compared to Luminal. We identified: 1) *BRCA1* inactivation, 2) *RBI* loss and *Cyclin E1* amplification, 3) high expression of *AKT3*, 4) *cMYC* amplification and high expression, and 5) a high frequency of *TP53* mutations (Figure 5A). An additional supervised analysis of a large, external multi-tumor type transcriptomic dataset (GSE2109) was performed where each TCGA breast tumor expression profile was compared via a correlation analysis to that of each tumor in the multi-tumor set. Basal-like breast cancers clearly showed high mRNA expression correlations with Serous Ovarian cancers, as well as with Lung Squamous carcinomas (Figure 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, cMYC, and FOXM1 regulatory hubs in both Ovarian and Basal-like cancers (Supplemental Figure 20C). The common findings of *TP53*, *RBI* and *BRCA1* loss, with *cMYC* 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 analogs, and taxanes, in breast Basal-like and Serous Ovarian cancers.

Given that most Basal-like cancers are TNBC, 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 tumors had a germline (n=12) and/or somatic (n=8) *BRCA1* or *BRCA2* variant, which suggests 1 in 5 Basal-like patients might benefit from PARP inhibitors and/or platinum compounds^{52,53}. The copy number landscape of Basal-like cancers showed multiple amplifications and deletions, some of which may provide therapeutic targets (Supplemental Table 6). Potential targets include losses of *PTEN* and *INPP4B*, both of which have been shown to sensitize cell lines to PI3K pathway inhibitors^{54,55}. Interestingly, many of the components of the PI3K and RAS-RAF-MEK pathway were amplified (but not typically mutated) in Basal-like cancers including *PIK3CA* (49%), *KRAS* (32%), *BRAF* (30%), and *HER1/EGFR* (23%). Other RTKs that are plausible drug targets and amplified in some Basal-like cancers include *FGFR1*, *FGFR2*, *IGFR1*, *c-KIT*, *MET* and *PDGFRA*. Lastly, the PARADIGM identification of high HIF1 α /ARNT pathway activity suggest that these malignancies might be susceptible to angiogenesis inhibitors and/or bioreductive drugs that become activated under hypoxic conditions.

Summary

The integrated molecular analyses of breast carcinomas we report significantly extends our knowledge base to produce a comprehensive catalog 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 time of data freeze; 466 samples (463 patients) were in common across 5/6 platforms (excluding RPPA) and 348 patients were in common on 6/6 platforms. Technology platforms used include: 1) gene expression DNA microarrays⁵¹, 2) DNA methylation arrays, 3) microRNA 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 (Supplemental Methods) which were included in a single analysis to define an overall subtype architecture. Additional integrated across platform computational analyses were preformed including PARADIGM³² and MEMO⁴⁰.

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 (DCC) (<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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Margi Sheth and Susan Lucas for administrative coordination of TCGA activities, and Chris 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 U.S. Government.

References

1. Paik S, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004; 351:2817–2826. [PubMed: 15591335]
2. van 't Veer LJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002; 415:530–536. [PubMed: 11823860]
3. Slamon DJ, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science.* 1987; 235:177–182. [PubMed: 3798106]
4. Chin K, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell.* 2006; 10:529–541. S1535-6108(06)00315-1 [pii]. 10.1016/j.ccr.2006.10.009 [PubMed: 17157792]
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 Chromosomes Cancer.* 2006; 45:1033–1040. [PubMed: 16897746]
6. Perou CM. Molecular stratification of triple-negative breast cancers. *Oncologist.* 2011; 16(Suppl 1): 61–70.10.1634/theoncologist.2011-S1-61 [PubMed: 21278442]
7. Sorlie T, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A.* 2003; 100:8418–8423. [PubMed: 12829800]
8. Foulkes WD, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst.* 2003; 95:1482–1485. [PubMed: 14519755]
9. Carey LA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama.* 2006; 295:2492–2502. [PubMed: 16757721]
10. Ding L, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* 2010; 464:999–1005. nature08989 [pii]. 10.1038/nature08989 [PubMed: 20393555]
11. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature.* 2009; 461:809–813.10.1038/nature08489 [PubMed: 19812674]
12. Shah SP, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature.* 2012.10.1038/nature10933
13. Perou CM, et al. Molecular portraits of human breast tumours. *Nature.* 2000; 406:747–752. [PubMed: 10963602]
14. Parker JS, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J Clin Oncol.* 2009 JCO.2008.18.1370 [pii]. 10.1200/JCO.2008.18.1370

15. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, *Condel. Am J Hum Genet.* 2011; 88:440–449.10.1016/j.ajhg.2011.03.004 [PubMed: 21457909]
16. Bamshad M, et al. Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat Genet.* 1997; 16:311–315.10.1038/ng0797-311 [PubMed: 9207801]
17. Li QY, et al. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet.* 1997; 15:21–29.10.1038/ng0197-21 [PubMed: 8988164]
18. The Cancer Genome Atlas Research Network et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008
19. Cheung LW, et al. High Frequency of PIK3R1 and PIK3R2 Mutations in Endometrial Cancer Elucidates a Novel Mechanism for Regulation of PTEN Protein Stability. *Cancer discovery.* 2011; 1:170–185.10.1158/2159-8290.CD-11-0039 [PubMed: 21984976]
20. Malcovati L, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood.* 2011; 118:6239–6246.10.1182/blood-2011-09-377275 [PubMed: 21998214]
21. Wang L, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med.* 2011; 365:2497–2506.10.1056/NEJMoa1109016 [PubMed: 22150006]
22. Ding L, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature.* 2008; 455:1069–1075.10.1038/nature07423 [PubMed: 18948947]
23. Johnson GL, Lapadat R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science.* 2002; 298:1911–1912.10.1126/science.1072682 [PubMed: 12471242]
24. Usary J, et al. Mutation of GATA3 in human breast tumors. *Oncogene.* 2004; 23:7669–7678. [PubMed: 15361840]
25. Walsh T, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A.* 2010; 107:12629–12633.10.1073/pnas.1007983107 [PubMed: 20616022]
26. Prat A, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* 2010; 12:R68. bcr2635 [pii]. 10.1186/bcr2635 [PubMed: 20813035]
27. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 2011; 39:D152–157.10.1093/nar/gkq1027 [PubMed: 21037258]
28. Weigman VJ, 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.* 2011; 10.1007/s10549-011-1846-y
29. Curtis C, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012; 10.1038/nature10983
30. Hennessy BT, 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. *Clinical proteomics.* 2010; 6:129–151.10.1007/s12014-010-9055-y [PubMed: 21691416]
31. Daub H, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell.* 2008; 31:438–448.10.1016/j.molcel.2008.07.007 [PubMed: 18691976]
32. Vaske CJ, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics.* 2010; 26:i237–245.10.1093/bioinformatics/btq182 [PubMed: 20529912]
33. Campbell IG, et al. Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res.* 2004; 64:7678–7681. [PubMed: 15520168]
34. Bachman KE, et al. The PIK3CA Gene is Mutated with High Frequency in Human Breast Cancers. *Cancer Biol Ther.* 2004; 3:772–775. [PubMed: 15254419]
35. Stemke-Hale K, et al. An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* 2008; 68:6084–6091. [PubMed: 18676830]

36. Creighton CJ, 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.* 2010; 12:R40.10.1186/bcr2594 [PubMed: 20569503]
37. Majumder PK, et al. mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nat Med.* 2004; 10:594–601.10.1038/nm1052 [PubMed: 15156201]
38. Saal LH, et al. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nat Genet.* 2008; 40:102–107. [PubMed: 18066063]
39. Wagner EF, Nebreda AR. Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat Rev Cancer.* 2009; 9:537–549.10.1038/nrc2694 [PubMed: 19629069]
40. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* 2012; 22:398–406.10.1101/gr.125567.111 [PubMed: 21908773]
41. Kannan K, et al. DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene.* 2001; 20:2225–2234. [PubMed: 11402317]
42. Troester MA, et al. Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer.* 2006; 6:276. [PubMed: 17150101]
43. Deisenroth C, Thorner AR, Enomoto T, Perou CM, Zhang Y. Mitochondrial Hep27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53. *Mol Cell Biol.* 2010; 30:3981–3993. MCB. 01284-09 [pii]. 10.1128/MCB.01284-09 [PubMed: 20547751]
44. Pei XH, et al. CDK inhibitor p18(INK4c) is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell.* 2009; 15:389–401. S1535-6108(09)00079-8 [pii]. 10.1016/j.ccr.2009.03.004 [PubMed: 19411068]
45. Chicas A, et al. Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell.* 2010; 17:376–387.10.1016/j.ccr.2010.01.023 [PubMed: 20385362]
46. Herschkowitz JI, He X, Fan C, Perou CM. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Res.* 2008; 10:R75. bcr2142 [pii]. 10.1186/bcr2142 [PubMed: 18782450]
47. Lara MF, et al. Gene profiling approaches help to define the specific functions of retinoblastoma family in epidermis. *Mol Carcinog.* 2008; 47:209–221.10.1002/mc.20376 [PubMed: 17932948]
48. Baselga J, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med.* 2012; 366:109–119.10.1056/NEJMoa1113216 [PubMed: 22149875]
49. 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.* 2010 41490 [pii]. 10.1172/JCI41490
50. 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.* 2009; 4:e6693.10.1371/journal.pone.0006693 [PubMed: 19690609]
51. Perou CM. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature.* 2011; 474:609–615.10.1038/nature10166 [PubMed: 21720365]
52. Audeh MW, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparibin patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet.* 2010; 376:245–251.10.1016/S0140-6736(10)60893-8 [PubMed: 20609468]
53. Fong PC, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med.* 2009; 361:123–134. NEJMoa0900212 [pii]. 10.1056/NEJMoa0900212 [PubMed: 19553641]
54. Fedele CG, et al. Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc Natl Acad Sci U S A.* 2010; 107:22231–22236.10.1073/pnas.1015245107 [PubMed: 21127264]
55. Gewinner C, et al. Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell.* 2009; 16:115–125.10.1016/j.ccr.2009.06.006 [PubMed: 19647222]

Appendix

Genome Sequencing Centers

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 Centers

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 and Women's Hospital and Harvard Medical School*– Lynda Chin^(9,11), Peter J. Park⁽¹²⁾, and 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 Brooker⁽¹⁸⁾, Junyuan Wu⁽¹⁵⁾, Anisha Gulabani⁽¹⁵⁾, Tom Bodenheimer⁽¹⁵⁾, Alan P.

¹The Genome Institute, Washington University, St Louis, Missouri 63108, USA

²Department of Genetics, Washington University, St Louis, Missouri 63110, USA

³Siteman Cancer Center, Washington University, St Louis, Missouri 63110, USA

⁴Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z, Canada

⁵The Broad Institute of MIT and Harvard, Cambridge, MA 02142

⁶Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, 02115

⁷Department of Medicine, Harvard Medical School, Boston, MA 02215

⁸Departments of Cancer Biology and Medical Oncology, and the Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA, 02115

⁹Department of Medical Oncology and the Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA, 02115

¹⁰Department of Pathology, Harvard Medical School, Boston, MA 02215

¹¹Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, MA 02115 USA

¹²The Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115 USA

¹³Department of Genetics, Harvard Medical School and Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115 USA

¹⁴Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

¹⁵Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

¹⁶Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

¹⁷Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

¹⁸Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, Chapel Hill, NC 27599 USA

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⁾, Jianjiang 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 and 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⁾,

¹⁹Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

²⁰USC Epigenome Center, University of Southern California, Los Angeles, CA 90033 USA

²¹Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD 21231 USA

²²Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030

²³Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030

²⁴Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030

²⁵Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030

²⁶The Eli and Edythe L. Broad Institute of Massachusetts Institute Of Technology and Harvard University, Cambridge, MA 02142 USA

²⁷Institute for Applied Cancer Science, Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX 77054

²⁸Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX 77054

²⁹Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115 USA

³⁰Informatics Program, Children's Hospital, Boston, MA 02115 USA

³¹Institute for Systems Biology, Seattle, WA 98109 USA

³²Tampere University of Technology, Tampere, Finland

³³Cancer Genomics Core Laboratory, M.D. Anderson Cancer Center, Houston, TX 77030 USA

³⁴Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

³⁵Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

³⁶Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

³⁷Oregon Health and Science University, 3181 SW Sam Jackson Park Rd, Portland OR 97239

³⁸Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

³⁹Kleberg Center for Molecular Markers, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

⁴⁰Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

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 Łańniak⁽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

⁴¹Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, CA 95064 USA

⁴²Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064 USA

⁴³Buck Institute for Research on Aging, Novato, CA 94945 USA

⁴⁴Center for Bioinformatics and Information Technology, National Cancer Institute, Rockville, MD

⁴⁵The Ohio State University College of Medicine, Department of Pathology, Columbus, OH 43205

⁴⁶The Ohio State University College of Medicine, Department Pediatrics, Columbus, OH 43205

⁴⁷The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205

⁴⁸ABS Inc. Indianapolis, IN 46204 USA

⁴⁹ABS Inc. Wilmington DE 19801 USA

⁵⁰Indiana University School of Medicine, Indianapolis, Indiana 46202 USA

⁵¹Helen F. Graham Cancer Center, Christiana Care, Newark, Delaware 19713 USA

⁵²Moscow City Clinical Oncology Dispensary #1 and the Central IHC Laboratory of the Moscow Health Department, Moscow, Russia

⁵³Russian Cancer Research Center, Moscow, Russia

⁵⁴Cureline, Inc., South San Francisco, CA, USA

⁵⁵Department of Surgery, Duke University Medical Center, Durham, NC 27710 USA

⁵⁶The Greater Poland Cancer Centre, Poznań, 61-866, Poland

⁵⁷Poznań University of Medical Sciences, Poznań, 61-701, Poland

⁵⁸ILSBio, LLC, Chestertown, MD 21620, USA

⁵⁹Ministry of Health, Hanoi, Vietnam

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 Gaudio⁽⁷⁵⁾, Wiam Bshara⁽⁷⁵⁾, *University of Miami* - Kelly Daily⁽⁷⁶⁾, Sophie C. Egea⁽⁷⁶⁾, Marc 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)

⁶⁰ILSBio LLC, Karachi, Pakistan

⁶¹Hue Central Hospital, Hue City, Vietnam

⁶²Stanford University Medical Center, Stanford, CA 94305, USA

⁶³Center for Minority Health Research, University of Texas, M.D. Anderson Cancer Center, Houston, TX 07703

⁶⁴National Cancer Institute, Hanoi, Vietnam

⁶⁵Ho Chi Minh City Cancer Center, Vietnam

⁶⁶Can Tho Cancer Center, Can Tho, Vietnam

⁶⁷International Genomics Consortium, Phoenix, AZ 85004 USA

⁶⁸Mayo Clinic, Rochester, MN 55905

⁶⁹Department of Surgery, Breast Service, Memorial Sloan-Kettering Cancer Center, New York, NY 10065

⁷⁰Department of Breast Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, TX 77030 USA

⁷¹University of California at San Francisco; San Francisco, CA 94143

⁷²Cancer Diagnostics; Nichols Institute, Quest Diagnostics; San Juan Capistrano, CA 92675

⁷³Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA 94115

⁷⁴UNC Tissue Procurement Facility, Department of Pathology, UNC Lineberger Cancer Center, Chapel Hill, NC 27599, USA

⁷⁵Department of Pathology, Roswell Park Cancer Institute, Buffalo, NY 14263 USA

⁷⁶Department of Pathology, University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, FL 33136, USA.

⁷⁷University of Pittsburgh, Pittsburgh, PA, 15213 USA

⁷⁸Magee-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh PA 15213 USA

⁷⁹Walter Reed National Military Medical Center, Bethesda, MD 20899-5600

⁸⁰Windber Research Institute, Windber, PA 15963

⁸¹MDR Global, LLC, Windber, PA 15963

⁸²Breast Cancer Program, Washington University, St. Louis, MO, USA

⁸³Department of Internal Medicine, Division of Oncology, Washington University, St. Louis, MO, USA

⁸⁴Institute for Genomics and Systems Biology, University of Chicago, Chicago IL

⁸⁵Center for Clinical Cancer Genetics, The University of Chicago, Chicago, IL 60637, USA

⁸⁶Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110

Data Coordination Center

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⁾

⁸⁷SRA International, 4300 Fair Lakes Court, Fairfax, VA 22033

⁸⁸The Cancer Genome Atlas Program Office, Center for Cancer Genomics, National Cancer Institute, Bethesda, MD

⁸⁹TCGA Consultant, Scimentis, LLC, Atlanta, GA

⁹⁰MLF Consulting, Arlington, MA 02474 USA

⁹¹National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892

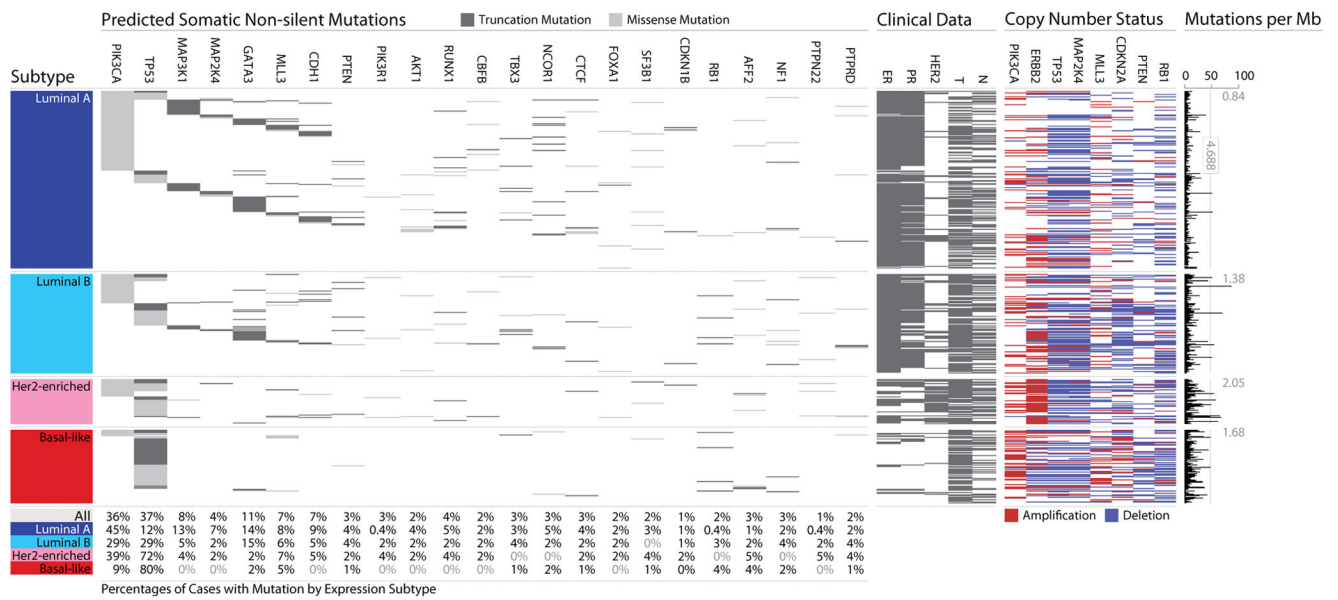


Figure 1. Significantly Mutated Genes (SMG) and correlations with genomic and clinical features

Tumor samples are grouped by mRNA-subtype: Luminal A (n=225), Luminal B (n=126), HER2E (n=57), and Basal-like (n=93). **Left:** Non-silent somatic mutation patterns and frequencies for SMGs. **Middle:** Clinical features: black, positive or T2-4; white, negative or T1; grey, NA or equivocal. **Right:** SMGs with frequent copy number amplifications (red) or deletions (blue). **Far Right:** Non-silent mutation rate per tumor (mutations per megabase, adjusted for coverage). Average mutation rate for each expression subtype is indicated. Hypermutated: mutation rates > 3 SD above the mean (> 4.688, indicated by grey line).

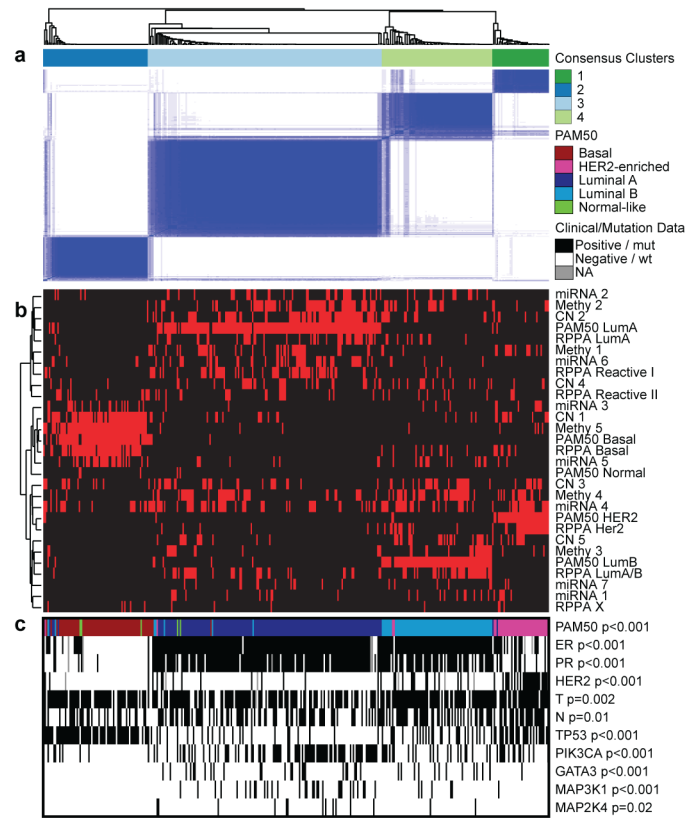


Figure 2. Coordinated analysis of breast cancer subtypes defined from five different genomic/proteomic platforms

a) Consensus clustering analysis of the subtypes identifies four major groups (samples, n=348). The blue and white heatmap displays sample consensus. b) Heatmap display of the subtypes defined independently by microRNAs, DNA methylation, copy number, PAM50 mRNA expression, and RPPA expression. Red bar indicates membership of a cluster type. c) Associations with molecular and clinical features. P-values were calculated using a Chi-square test.

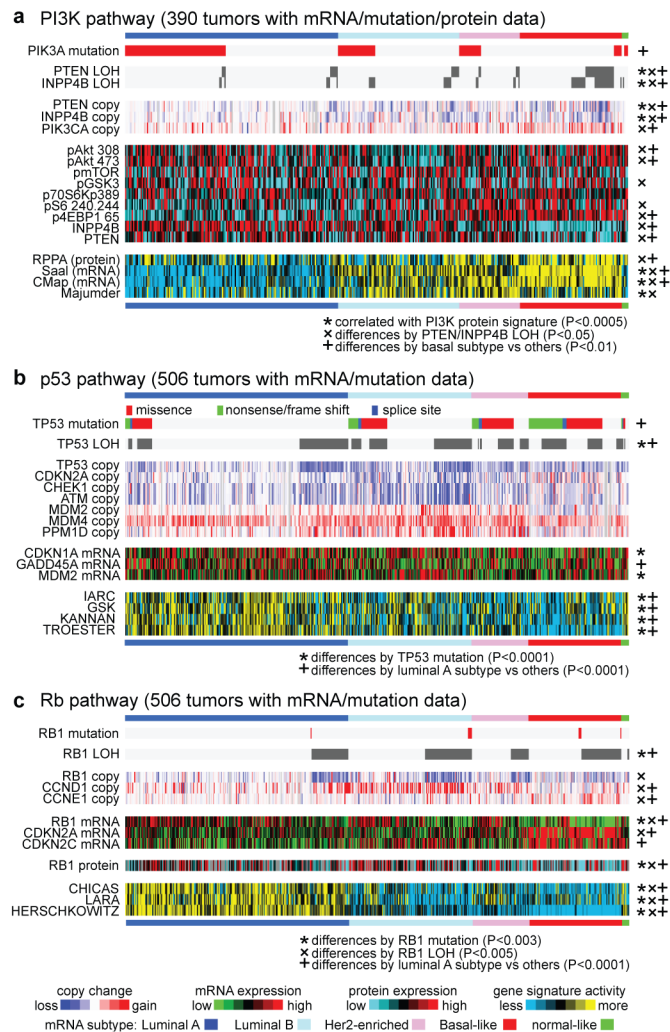


Figure 3. Integrated analysis of the PI3K, TP53, and RB1 pathways

Breast cancer subtypes differ by genetic and genomic targeting events, with corresponding effects on pathway activity. For a) PI3K, b) TP53 and c) RB1 pathways, key genes were selected using prior biological knowledge. Multiple mRNA expression signatures for a given pathway were defined (details in Supplemental Methods; *PI3K:Saal*, PTEN loss in human breast tumors; *PI3K:CMAP*, PI3K/mTOR inhibitor treatment *in vitro*; *PI3K:Majumder*, Akt over-expression in mouse model; *p53:IARC*, expert-curated p53 targets; *p53:GSK*, TP53 mutant versus wild-type cell lines; *p53:KANNAN*, p53 over-expression *in vitro*; *p53:TROESTER*, TP53 knockdown *in vitro*; *Rb:CHICAS*, RB1 mouse knockout versus wild-type; *Rb:LARA*, RB1 knockdown *in vitro*; *Rb:HERSCHKOWITZ*, RB1 LOH in human breast tumors) and applied to the gene expression data, in order to score each tumor for relative signature activity (yellow: more active). The PI3K panel includes a protein-based (RPPA) proteomic signature. Tumors were ordered first by mRNA-subtype, though specific ordering differs between the panels. P-values were calculated by a Pearson's correlation or a Chi-squared test.

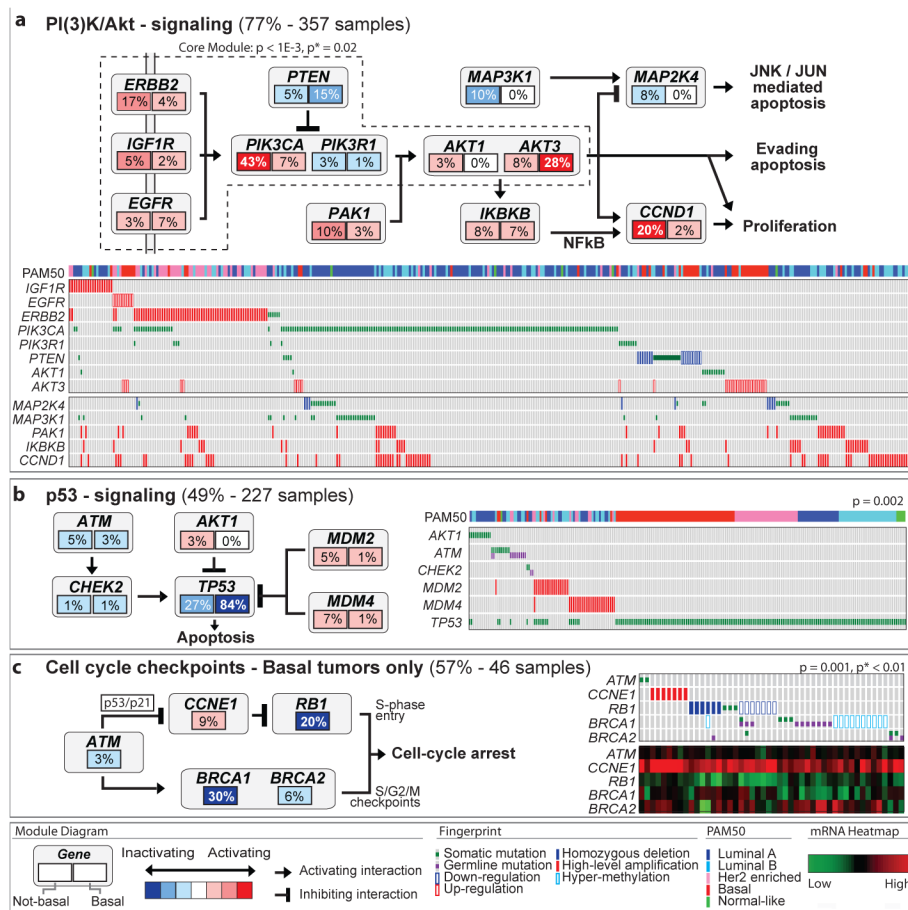


Figure 4. Mutual Exclusivity Modules in Cancer (MEMo) analysis

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/PI3K and p38/JNK1 signaling pathways and whose core was the top-scoring module. b) MEMo identified alterations to TP53 signaling 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 heatmap is below the fingerprint to show expression levels.

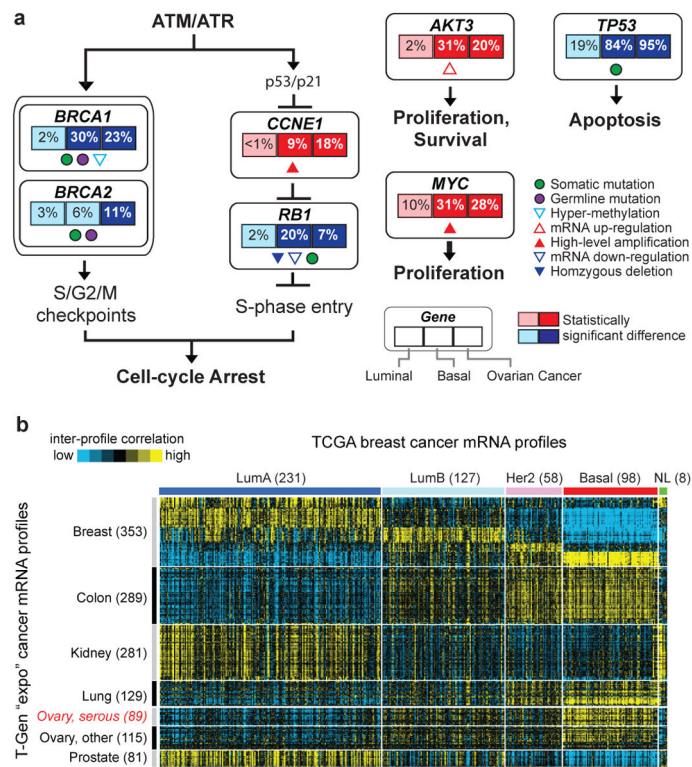


Figure 5. Comparison of Breast and Serous Ovarian carcinomas

a) Significantly enriched genomic alterations identified by comparing Basal-like or Serous Ovarian tumors to Luminal cancers. b) Inter-sample correlations (yellow: positive) between gene transcription profiles of breast tumors (columns; TCGA data, arranged by subtype) and profiles of cancers from various tissues of origin (rows; external "TGEN expO" dataset, GSE2109) including Ovarian cancers.

Table 1

Summary of disease subtypes findings highlighting some of the dominant genomic, clinical, and proteomic features.

| | Luminal A | Luminal B | Basal-like | HER2E |
|---------------------|---|---|---|---|
| % ER+/HER2- | 87% | 82% | 10% | 20% |
| % HER2+ | 7% | 15% | 2% | 68% |
| % TNBC | 2% | 1% | 80% | 9% |
| p53 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 amps 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 ERBB2 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 | | Hyper-methylated phenotype for subset | Hypo-methylated | |
| Protein Expression | High Estrogen-signaling High cMYB RPPA reactive subtypes | Less Estrogen-signaling High FOXM1 and cMYC RPPA reactive subtypes | High expression of DNA repair proteins, PTEN and INPP4B loss signature (p-AKT) | High protein and phospho-protein expression of HER1 and HER2 |

Percentages are based on 466 tumor overlap list