Published in final edited form as:

J Mammary Gland Biol Neoplasia. 2017 March ; 22(1): 71–84. doi:10.1007/s10911-017-9374-y.

Mouse models of breast cancer share amplification and deletion events with human breast cancer

Jonathan Rennhack¹, Briana To¹, Harrison Wermuth¹, and Eran R. Andrechek^{1,2}

¹Department of Physiology, Michigan State University

Abstract

Breast tumor heterogeneity has been well documented through the use of multiplatform –omic studies in human tumors. However, there is no integrative database to capture the heterogeneity within mouse models of breast cancer. This project identifies genomic copy number alterations (CNAs) in 600 tumors across 27 major mouse models of breast cancer through the application of a predictive algorithm to publicly available gene expression data. It was found that despite the presence of strong oncogenic drivers in most mouse models, CNAs are extremely common but heterogeneous both between models and within models. Many mouse CNA events are largely conserved in human tumors and in the mouse we show that they are associated with secondary tumor characteristics such as tumor histology, metastasis, as well as enhanced oncogenic signaling. These data serve as an important resource in guiding investigators when choosing a mouse model to understand the gene copy number changes relevant to human breast cancer.

Keywords

copy number variation; mouse model; breast cancer; gene expression; metastasis

Introduction

Genomic instability, including point mutations, translocation, and gene copy number alteration in key oncogenic signaling genes, is an underlying driver of breast cancer development and progression. Gene copy number changes, containing amplifications such as *HER2* (1) and *MYC* (2) or deletion events such as *PTEN* (3), are key biomarkers of tumor onset (4–7), histology (8), metastatic potential (9), and treatment response (10, 11). The amplification of *HER2* in 20–30% of breast cancer patients results in significantly more aggressive tumors and is an important prognostic marker in a patient's ability to respond to anti-HER2 therapy such as trastuzumab. Despite the success of HER2 therapy, the majority

²To whom correspondence should be addressed. Eran R. Andrechek, Department of Physiology, Michigan State University, 2194 BPS Building, 567 Wilson Road, East Lansing, MI, 48824, (517)884-5042 Office, (517)884-5020 Lab, (517)355-5125 Fax.

Conflicts of Interest

The authors declare they have no conflict of interest

Authors' contributions

JR and EA collaborated on the study conception, design, and interpretation of results. BT provided histological coordination with CNA. HW provided the ACE analysis for many mouse models. JR performed all other experiments and drafted the manuscript. All authors have critically read, edited, and approved the final version of the manuscript.

of patients with the amplification event will have primary or acquired resistance to trastuzumab (11), indicating potential heterogeneity in these tumors.

To investigate tumor heterogeneity, large multiplatform studies such as The Cancer Genome Atlas (TCGA) (12) and Metabric (13) projects have begun to integrate transcriptional data and genomic data, as well other data platforms. Traditionally, breast cancer diversity has been classified at the transcriptional level into six basic subtypes: basal, luminal A, luminal B, HER2 positive, claudin-low, and normal like (14, 15). The integration of gene copy number showed that there are a number of classical gene copy number alterations that are associated with each tumor subtype. For instance, this revealed that *MYC* amplification occurs in all breast cancer subtypes, but *MYC* is only transcriptionally active in the basal subtype (12). This study underscores the importance of integration of multiple platforms to understand tumor heterogeneity.

In order to understand the function of oncogenic drivers, research has employed mouse models. There are a variety of methods to induce breast cancer in a mouse model. These range from tissue specific overexpression using promoters including mouse mammary tumor virus (MMTV) or Whey acidic protein (WAP) promoters to drive oncogene expression (16–19), to conditional knockouts of tumor suppressive genes through the use of a tissue specific Cre (20–22) or inducible system (23, 24) and carcinogen induced model such as DMBA treatment. Furthermore, models have been used to investigate particular aspects of tumor development such as metastasis via MMTV-PyMT (25) or genomic instability with loss of p53 (26). Given these varied methods and drivers of tumor formation, the transcriptional program in each model would be expected to be unique. Importantly the recent advent of patient derived xenografts (PDX) models has given a new option. These models have been shown to reflect their human counterpart at a genomic (27) and transcriptional level; however, other common mouse models of breast cancer have not been described in such a manner.

Recent work has captured the gene expression diversity between models and within tumors of the same model (28–30). However, these works do not describe multiple levels of genomic diversity in mouse models like the TCGA and Metabric projects do in human tumors. Largely this is due to a lack of multi-platform omic studies present for the mouse models. Small scale studies that integrate CNA with expression data across species have identified a unique CNA in Basal like breast cancer (31). The lack of such profiling on a large scale leaves researchers relatively uninformed about genomic changes present in a mouse tumor when choosing a mouse model which is representative of a specific subtype of human breast cancer.

Here we describe a large scale investigation of copy number changes in 600 tumors across 27 mouse models of breast cancer for the use of the ACE algorithm (32). In short, the ACE algorithm predicts CNA from gene expression data through the use of a weighted mean of gene expression across a given genomic region. Due to its reliance on consistent regulation across an entire genomic region it has been shown to accurately predict copy number variants and has been shown to have consistent results to traditional genetic predictors of gene copy number tumors (32). This predicted CNA across our dataset demonstrated wide

heterogeneity across mouse models of breast cancer. Interestingly, consistent CNA changes were noted in the microacinar histological subtypes of breast cancer indicating a role in copy number changes and a tumor's histological phenotype. Moreover, in an important observation we noted that CNA was associated with breast cancer metastasis and enhanced oncogenic signaling in both mouse models and human breast cancer.

Results

Identification of gene copy number alteration in mouse models of breast cancer

A large number of mouse model tumors have been examined by microarray for gene expression but few have been assessed for genome wide copy number alteration. To computationally predict CNA from gene expression data, we applied the ACE algorithm (32). The ACE algorithm is shown to be consistent with traditional means of determining CNA utilized by the TCGA breast cancer study. To validate the algorithm, we used the entire TCGA breast cancer dataset in which both gene expression and copy number data is available. A random selection of three samples from the dataset show a high degree of similarity between the ACE calls and the TCGA CNA calls (Figure 1A) (12, 33). Application of the algorithm across the entire dataset shows a false positive rate of 25.4% and 24.3% for amplifications and deletions respectively for the ACE algorithm (Figure 1B). However, the false negative rate was higher at greater than 90% (Figure 1B). This includes both amplification of whole arms of the chromosome and very small amplification events. For the false negatives, the algorithm also shows a dependence upon gene expression being coordinated with CNA. Importantly, in the case of amplification events with a high transcriptional impact, noted by a z-score of greater than 4, we found a success rate of 31.1% of TCGA events called by the ACE algorithm.

We have also shown the translational effects of copy number gains and losses in human tumors through investigation of Reverse Phase Protein Array (RPPA) data associated with *EGFR* (Supplemental Figure 1A) and *FOXO3* (Supplemental Figure 1B) amplification events. This analysis shows that in both *EGFR* and *FOXO3* the protein level is directly correlated with gene copy number. This validation of the ACE software across the TCGA dataset shows that the events predicted in this manuscript are an understatement of the events in the tumor. However, the false low false positive rate shows that the events called in the manuscript can be used for predictive purposes and begin to show the copy number profile in mouse models of breast cancer.

To investigate the presence of copy number alterations in mouse models we applied the ACE algorithm to gene expression data from a normal mammary gland from an FVB Wildtype mouse (Figure 1C) and an MMTV-Neu derived tumor from the same background (Figure 1D). As expected, no CNA was identified in the control mammary gland. In contrast, the MMTV-Neu sample is characterized by a large amplification event on chromosome 3 as well as a large deletion on chromosome 4. This deletion is consistent with previously published findings of chromosome 4 loss in MMTV-Neu mouse models (34). In addition to these major CNA events, there are a number of smaller CNAs throughout of the genome of the sample.

We then hypothesized as a further check that unstable models would have significantly more CNAs than oncogene induced models. To investigate this hypothesis, we tested for CNA in multiple samples from unique mouse models of breast cancer. The ACE algorithm was applied to tumor samples from MMTV-Myc, MMTV-PyMT, MMTV-Neu, TAG, and DMBA treated models derived from the FVB/NJ mouse background (Figure 1E). ACE analysis showed genomic stability in the MMTV-Neu driven mouse models. This model had significantly fewer amplification or deletion events than more classically unstable models such as the TAG ($p < .05$) and DMBA ($p < .05$). This is mirrored in human cancer where certain tumors such as Basal tumors are shown to be more unstable than other subtypes especially Luminal A (35). It was also noted that mice with the same oncogenic initiation event such as PyMT had a difference in copy number based upon the background of the mouse model (Supplemental Figure 2). We noted that the FVB background was the most unstable when compared to the AKXD background in PyMT driven tumors and the Balb/C background in the TAG or various p53 driven models.

Conservation of CNA Variability in Mouse Models

To investigate the extent of CNA in mouse models and to determine if there were common alterations across various oncogenic driver genes, expression data was downloaded from a previously assembled mouse model database (28). Specifically, for copy number variability we used 600 tumor samples from 27 mouse models that had been analyzed through the use of Affymetrix microarrays (Table 1). ACE analysis was run and the percent of samples, regardless of model, amplified or deleted with CNA at a specific locus across the genome was calculated (Figure 2A).

This data shows the vast majority of genes across the genome are amplified or deleted in less than 5% of the total samples with a few distinct regions being amplified or deleted in a larger fraction of samples. However, we did identify regions of instability that were conserved across models. We identified a number of genes that were both amplified and deleted in greater than 10% of mouse models. Specifically, we identified *Gsn* is amplified 10.9% of samples and deleted in 11.4% of samples. Other genes that were amplified and deleted at a high level included *Cct4*, *Hnmpab*, *Cp*, *Cklf*, *Cenpo*, and *Dnm2*. These genes are all located at regions previously described in the mouse genome to be unstable (36). The percentage of amplification or deletion for all genes can be found in Supplementary Table 1.

Given the extensive heterogeneity in breast cancer, we sought to test the hypothesis that heterogeneity was present at the level of gene copy number within individual mouse models of cancer. The extent of heterogeneity of CNAs within a tumor model was analyzed by examining the fraction of mice within a model with a given amplification or deletion event at a particular locus (Figure 2B, Table S1). This analysis revealed a large degree of heterogeneity within models with most models having the majority of loci amplified or deleted in less than 50% of the sample within a model. This is despite that fact that many tumor samples are driven by the same oncogenic driver and are biological replicates. Some of the genes that were amplified in greater than 50% of samples within a given tumor model represent key genes in tumor development, progression and metastasis including well known genes such as *Cdkn2*, *Mmp23*, *Sumo2*, and *Adcy33*. Interestingly, conserved CNA events

were not seen to span models, reinforcing the genomic diversity both within each model system and between the model systems. In addition, we noted some models with more copy number events, such as the p53 induced models.

These CNA changes were then divided into amplification events (Supplemental Figure 3A) or deletion events (Supplemental Figure 3B) to reflect the copy number diversity in each model. This revealed that mouse tumor models largely fall into three categories. First, we observed unstable models with a high degree of amplification or deletion in a large number of genes but with low levels of conservation, this including many models with a *p53* mutation. Secondly, we noted models that are relatively stable with no amplification or deletion at the vast majority of genes, including models such as MMTV-PyMT. Lastly there are models with a few highly conserved amplification or deletion events. These conserved events were noted in more than 25% of samples in lines such as the *ErbB2* knock-in model or the WAP TAG model.

Conserved Role of CNAs in Human and Mouse Tumors

A key feature of the mouse tumors is their ability to model human cancer. To test the hypothesis that there were conserved CNAs in both species, we began by identifying the fraction of tumors within each model with an amplification or a deletion at genes prone to copy number alteration in human breast cancer such as *ERBB2*, *MYC*, *PTEN*, *RB*, and others identified in the TCGA study. Driver genes were found to be amplified or deleted in specific mouse models. Specifically, this occurred in the BRCA/p53 modified models, which have a fraction of samples with amplification in common oncogenes such as *CCNE* or deletion of common tumor suppressors like *CMTM3*.

To identify genes that were amplified or deleted at a high level in both mouse and human that are not traditional drivers of tumorigenesis, we used unsupervised hierarchical clustering of CNAs in human tumors of various subtypes and mouse models of breast cancer (Supplemental Figure 4A). As expected, human tumors and mouse models showed diverse copy number profiles and in general showed similarity in copy number profiles. This was illustrated through co-clustering in an unsupervised hierarchical clustering of gene loci amplified or deleted in more than 5% of mouse and human tumors (Figure 3A). Through this analysis we identified three tightly clustered groups of human and mouse samples. The sub-cluster indicated by the purple portion of the dendrogram was largely dominated by human Luminal A and normal like tumors while the yellow and green portions of the dendrogram had clusters characterized by the presence of Luminal B tumors. Largely absent in this analysis were the HER2 positive tumors. This finding indicates the lack of mouse tumor models with the HER2 amplification event or associated copy number changes. Of interest we noted that a fraction of MMTV-PyMT samples were present in each of the clusters, indicating considerable diversity within the MMTV-PyMT tumor model from a gene copy number perspective. To show that within these clusters there is similarity between mouse and human samples, we showed a significant increase in the Jaccard index (a similarity metric of mouse to human samples) within the purple cluster, and a low Jaccard index between the mouse samples of the purple cluster and the human samples of the other two

defined clusters (Figure 3B). This shows that each cluster of tumors has a significantly different copy number profile.

While we noted model to model heterogeneity, our previous analysis revealed within model heterogeneity. It was hypothesized that the heterogeneity was due to differences in histological subtypes. To test this, unsupervised hierarchical clustering of CNA events from the MMTV-Myc tumor model by the top 4118 commonly amplified or deleted genes, filtered by standard deviation of neighborhood score, was performed (Figure 4A). The major clusters demonstrated that distinct copy number changes are associated specific histological subtypes within the *Myc* driven model. Specifically, there is a cluster enriched for the microacinary subtype. The microacinary subtype shows the majority of samples containing amplification of many genes located on chromosomes 11 and 15. The EMT subtype is characterized by having few amplification or deletion events. Tumors of this histological subtype have previously been noted to have activating mutations in *Kras* that contribute to the EMT histological subtype. These activating mutations may result in a reduced requirement for other copy number events.

The genes amplified in mice with a microacinary histological subtype, located on mouse chromosome 11, are also conserved in humans. Specifically, we found a region of fourteen genes which mapped to chromosome 17q25.1 in humans. These genes are amplified shown to be amplified in a subset of breast tumors as identified by cBio portal. An assessment of mouse (Figure 4B), and human (Figure 4C) tumors, from the MMTV-Myc mouse model and TCGA breast cancer dataset respectively, showed conserved histology across species.

Human tumor samples were divided into those containing the microacinary specific genes through the use of a 14 gene signature (Figure 4D). Tumors with at least two of the 14 genes amplified were considered to be in the amplified subgroup. This produced a subgroup for 28 human tumors and was compared against 30 randomly chosen tumors that contained none of the 14 amplified microacinary associated genes. Histology for each of the groups of samples was determined and it was found that the microacinary associated gene amplified subgroup contained an enrichment of tumors with a microacinary histology (Figure 4E). This indicated a conserved role in gene copy number across mouse and human breast cancer in determining tumor histology specifically with respect to the microacinary subtype.

To investigate the role of the CNA on tumor progression, we examined tumor metastasis by integrating gene expression data with gene copy number data. Specifically we used a previously identified lung metastasis gene signature (37) and correlated gene copy number events with the sample's metastasis score through Spearman's Rank correlation (Figure 5A) to reveal amplification and deletion events associated with highly metastatic samples. This was applied to the human TCGA breast dataset (Figure 5B – top) as well as the mouse model dataset (Figure 5B – bottom). When differences in gene location were taken into account, there were 132 gene copy number alterations highly correlated with metastasis that were conserved in the TCGA breast cancer and mouse datasets (Figure 5B and Table S2). To provide validation of the 132 genes and their association with tumor metastasis we leveraged the KM-plot human dataset to show that over express or decreased expression of amplified or deleted genes was associated with worse distant metastasis free survival. We showed that

55% of these regions had significantly increased or decreased metastasis free survival depending upon the transcript level of the gene. To further investigate the metastasis related genes we used the Metabric data with associated overall survival data. Of the samples with an identified amplification or deletion event in a predicted metastasis region, we showed 28% of the events resulted in a decrease in overall survival of the patients. A closer examination of mouse chromosome 3 (Figure 5C) revealed a number of genes in the 3F region where amplification is associated with a high metastasis score. It was seen that some regions, such as the 3F region, associated with the metastasis signature. It was seen that chromosome 3F amplification was also associated with high RAS activity revealing a potential mechanisms of metastasis (Figure 5C).

To test the role of chromosome 3F on metastasis, we examined mouse tumor samples where metastasis data and pathway activity predictions were available. In particular, we used the MMTV-Myc, MMTV-Neu, and MMTV-PyMT models. As predicted those samples with the 3F amplification had much higher predicted *Ras* activity and number of lung metastases than those with a deletion. (Figure 5D). The 3F amplification event is also conserved in human Luminal A tumors. When tumors were split on the basis of amplification of the analogous human region it was seen that they exhibited higher *Ras* pathway activity (Figure 5E) and had worse metastasis free survival (Figure 5f).

Given that CNA was associated with metastatic progression we then hypothesized that CNA would also impact key cell signaling pathways. To test this hypothesis we examined the role of CNAs on major oncogenic pathways including *BCAT*, *SRC*, *E2F*, and others (38, 39). This experiment then used the same workflow to coordinate CNA and pathway activation status as was used to coordinate amplification and deletion events with the tumor metastasis signature (Supplemental Figure 5A). This analysis revealed amplification and deletion regions associated with each major oncogenic signature. The regulation of signaling pathways can occur through amplifying key genes within the signaling pathway. An example of this was observed when specific amplified genes associated with high *AKT* activity located on chromosome 4 or the specific amplified genes associated located on chromosome 14 associated with high *E2F2* activity were tested. When these genes are displayed in an interaction network, the vast majority of the genes can be found to be located either up or downstream of their respective key signaling protein such as or RB/E2F2 (Supplemental Figure 5B). This suggests the chromosome 14 region is associated with Rb/E2F signaling.

Discussion

Here we have described the copy number alteration across the genome of 27 mouse models of breast cancer. This has been completed through the use of an algorithm (ACE) to infer gene copy number profile from gene expression data. The ACE algorithm was identified to have a high rate of false negatives and a relatively low rate of false positive calls. When the algorithm was run across the TCGA dataset, it was seen to have a moderate rate of concurrence with the TCGA copy number calls. Due to this, it is important to note the predictive nature of this database. While the copy number calls found in this dataset have not been validated using traditional means, the dataset begins to identify potential copy number variants in the mouse models of breast cancer. This is an important step in understanding

tumorigenesis in these models specifically from a copy number point of view until a more robust and accurate profiling of the tumors can be completed.

The copy number profiles have been examined in a number of ways to classify inter and intra model heterogeneity as well as the similarities between copy number profiles in mouse models and human breast cancer. This study makes important contributions in understanding CNA in mouse models. Beyond this the CNAs are profiled for their contribution to tumor progression and the conservation of this role in human tumors.

Despite the presence of strong oncogenic signals, gene copy number alterations are still extremely common in mouse models of breast cancer. Common human drivers of breast cancer such as *HER2*, *MYC* or *PTEN* were not observed to be amplified or deleted at a high level across mouse models. This is unsurprising due to the lack of selective pressure for CNAs in these oncogenes or tumor suppressors because of the presence of a strong oncogenic signal. The exceptions to this are the p53/BRCA induced models which do not have a strong oncogenic signal but instead induce genomic instability. Amplification or deletion events in common human oncogenes are more frequent in these models.

A key finding of this manuscript is the heterogeneity of copy number alterations both within a model and between models. The within model heterogeneity is surprising due to the fact that each tumor is a biological replicate with the same driving oncogenic event. We identified that most events that occur within a given tumor model are not shared among even 50% of tumors from that same model. This finding underscores the importance of gene expression and genomic characterization of tumor studies when dealing with mouse models due to the inherent genomic variability. It further emphasizes the need for a large enough cohort to capture the heterogeneity of all tumor models.

During preparation of this manuscript, a complementary study was published examining CNA in mouse models of breast cancer (40). This publication uses a different algorithm to predict CNAs, one that predicts resolution on the whole chromosome scale while the ACE algorithm provides finer resolution. Due to the predictive nature of defining gene amplification events from gene expression data, we believe that it is important to compare their manuscript with the data herein. This demonstrates that multiple algorithms call the same dataset with overlapping findings, resulting in a comprehensive view of CNA in mouse model tumors. The two manuscripts agree on a number of findings including the stability of mouse models with rapid latency, the p53 KO model being the most unstable, within model heterogeneity, and the association of CNA changes with the microacinar subtype. The increased precision of the ACE method has allowed us to identify small focal events in many of the models including PyMT that the published paper did not uncover. Indeed, the data we present here allows one to search for a mouse model with amplification or deletion of particular genes. We have also leveraged human data through the use of the TCGA, Metabric, and KMplotter datasets to provide a comprehensive comparison of mouse models and the five main subtypes of human breast cancer tumors. In addition, we have also shown the conservation of regions between the two species to predict a number of new metastasis related copy number changes.

We noted that the amplification or deletion events are associated with secondary tumor characteristics such as tumor histology, enhanced oncogenic signaling, and tumor metastasis. Specifically, we observed unique copy number profiles for the microacinar tumor histology including the amplification of fourteen genes on chromosome 17q25.1. However other histological subtypes did not have characteristic copy number profiles. Surprisingly, we noted that EMT tumors were stable in regards to copy number change, likely due to activation of *Kras* in MYC tumors (23, 41, 42). The lack of pattern of amplification or deletion of other histological subtypes indicates that there are other factors such as point mutations or transcriptional changes associate with these subtypes. This can also be said for oncogenic signaling pathways and tumor metastasis. While CNAs contribute to each of these, there are also contributions of single nucleotide variants (SNVs) and transcriptional changes. For this reason, it is important to integrate multiple platforms to understand tumor heterogeneity.

There is conservation between mouse and human subtypes in regards to tumor metastasis and oncogenic signaling. 132 genes that were amplified or deleted in mouse and human contributed to increased metastasis. Furthermore, these genes were located in the same oncogenic signaling pathways indicating conserved mechanisms of metastasis in human and mouse.

When comparing heterogeneity of breast cancer, we found a large degree of heterogeneity both between models and within specific models. Given these findings, it is therefore critical to understand copy number profile when choosing a strain to model human breast cancer. For example, if one is interested in the *HER2* oncogene there are a number of mouse models including the MMTV-Neu (16, 25), *ErbB2* Knock-in (20), NDL (43) and others with conditional activation (24). Each of these models has completely different CNA and transcriptional profiles leading to different oncogenic signaling and subsequently different tumor properties.

This heterogeneity also exists in other common models such as MMTV-Myc. This strain has previously been identified to be heterogeneous from a transcriptional viewpoint (41) and therefore it is unsurprising that it is also heterogeneous from a copy number standpoint. Due to the heterogeneity present at a gene expression and copy number level, investigators must take care when choosing tumor models of breast cancer to ensure that the chosen model reflects all aspects of the human breast cancer subtype they wish to model. We have also noted strain specific differences for some of the models. It was seen that the FVB model was found to be more unstable when compared to tumors derived from other backgrounds. This finding emphasizes the importance of researchers understanding the background of their mouse strain when choosing mouse models for their study.

Projects such as TCGA have profiled human tumors at multiple levels. This allows researchers to stratify human tumors by gene expression, copy number profile, as well as SNVs and epigenetic markings to find a tumor population that is relevant for their study. However, there is not a mouse model equivalent to this dataset, so researchers are unable to choose mouse models which represent their specific tumor subtype at multiple levels. Recent studies such as this and others have begun to make strides in this area by profiling tumors at

a CNA and expression levels. However, there is still a need to continue to profile mouse models through the use of whole genome sequencing as well as epigenetic markings. This information needs to be available to researchers in order to design studies that accurately represent the human subtypes of breast cancer.

This study clearly illustrates the importance of gene copy number alterations in tumor progression even in the presence of strong oncogenic drivers. Many mouse models contain a high degree of gene copy number alterations. These copy number alterations are highly heterogeneous both between models and within a model of breast cancer. Despite this heterogeneity, it was seen that the CNAs found in mice are conserved in humans. Conserved variants were associated with tumor progression and potentially play a role in enhanced oncogenic signaling, histological appearance, and the tumor's metastatic potential in both human and mouse tumors.

Beyond the profiling of mouse tumors and the conserved roles of CNAs in mouse and human tumors this study has a broader impact on the field of cancer research. It, when used in combination with gene expression studies, begins to create a comprehensive molecular portrait of tumors derived from mouse models of breast cancer. These studies could be significantly enhanced if outcome, pathology, metastasis and other clinical data was included when publishing tumor data from mouse models. However, this current study provides an essential resource to researchers to contemplate as they choose a model system to mimic a specific subtype of human breast cancer.

Materials and Methods

Dataset and ACE analysis

A comprehensive mouse dataset was downloaded and assembled as previous described (28) including GSE15263, GSE3165, GSE37954, GSE32152, GSE10450, GSE22406, GSE42533, GSE15904, GSE8836, GSE27101, GSE30864, GSE20416, GSE10193, GSE23938, GSE15119, GSE16110, GSE25488, GSE21444, GSE8828, E-TABM-684. ACE analysis was run as previously described (32) comparing each individual sample to a wildtype mouse of the same strain (FVB/NJ = GSE25488, Balb/C = GSE21444, C57BL/6 = GSE14753) with a significance threshold p and q value of .05 with any size of the event.

Z-score calculation

The microarray based expression data was downloaded from the TCGA dataset. Z score was calculated for each gene in each sample and each event was classified based on the Z-score for validation analysis.

Human Dataset and Analysis

The TCGA breast cancer and KMplot.com datasets were used for human copy number analysis (12) and validation of results. Specific tumor breast cancer subtype and copy number calls were used from the TCGA dataset (12, 33). To run ACE the gene symbols were replaced with their Affymetrix U133A_2 probe ID. This was queried using the cBio portal visualization tool. Distant metastasis free survival results were obtained using the

KMplot.com dataset (44). ACE analysis for the human analysis compared expression to normal HMEC gene expression (n=10) data gathered from GSE24468.

Mouse Metastasis Dataset

A dataset with known lung metastasis from MMTV-Neu (45), MMTV-Myc (46, 47), and MMTV-PyMT (48) was compiled for metastasis free survival of mouse models.

Mouse and human gene location conversion

Locations of mouse and human genes were taken from the Affymetrix array annotation files from mouse 430A_2 and human U133A_2 array. These locations were merged by common gene symbol to provide a conversion table between the two species for the location of a particular gene.

Clustering of human and mouse tumors

ACE analysis was performed as previously described on the TCGA breast cancer human dataset as well as the mouse dataset. Significant CNAs were mapped onto the mouse genome for clustering. For human to mouse comparisons genes were filtered to those genes that were amplified or deleted in at least 5% of human and mouse tumors (n=594). Unsupervised hierarchical clustering was performed using cluster 3.0 and Java Tree View. For tumor histology the MMTV-Myc dataset with histological annotations, GSE15904, was used. To cluster this dataset we filtered the genes to 4118 genes through the use of standard deviation of the neighborhood score. This removed all genes that were unaltered across the dataset. For all clustering analysis Euclidian distance, complete linkage was used for the similarity metric and clustering method respectively.

Jaccard Index

Jaccard index was calculated between clusters through use of the R package “sets” through use of the similarity function.

Mouse and human histological comparisons

Histological annotations were analyzed from a group of MMTV-Myc mouse tumors (41), GSE15904, as well as the human TCGA tumors (12). We identified the genes within mouse chromosome 11 which were also amplified in a subset of human tumors using cbio portal. These fourteen identified genes mapped to the 17q25.1 region in humans and are referred to as the microacinar associated event. For overrepresentation analysis we compared the number of human tumors with the microacinar subtype from a group with the microacinar associated amplification event found in mice against a random set of tumors not containing that amplification event through the use of a 2×2 contingency table.

Oncogenic Signature application

Predefined oncogenic signatures were applied to the dataset. Briefly, the training data was merged with the full dataset and batch effects removed through the use of COMBAT. These samples were then subjected to binary regression analysis with a predefined gene list and conditions for each individual signature (38, 39, 41, 49, 50).

Coordination of CNA with oncogenic signature

Oncogenic signatures (38, 39, 41, 50) and lung metastasis signatures (37) were applied to mouse and human datasets as previously described. These scores were coordinated to neighborhood score through the use of a Spearman rank correlation applied through R. A significance threshold of $P < .01$ was applied and the results were visualized using MATLAB.

Gene network interaction

Interaction networks were visualized through the use of STRING-DB (51). Input nodes were those genes significantly correlated with the particular pathway in a specific region as well as key signaling proteins for the pathway (Rb/E2F2). Twenty additional white nodes were added to complete the network.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the members of the Andrechek laboratory for helpful discussions.

Funding

This work was supported with NIH R01CA160514 to E.R.A and 1F99CA212221-01 to J.R..

This work was supported with NIH R01CA160514 to E.R.A. and NIH F99CA212221 to J.R

References

- Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science*. 1989; 244(4905):707–12. [PubMed: 2470152]
- Escot C, Theillet C, Lidereau R, Spyrtos F, Champeme MH, Gest J, et al. Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas. *Proc Natl Acad Sci U S A*. 1986; 83(13):4834–8. [PubMed: 3014513]
- Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*. 1997; 275(5308):1943–7. [PubMed: 9072974]
- Deming SL, Nass SJ, Dickson RB, Trock BJ. C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. *Br J Cancer*. 2000; 83(12):1688–95. [PubMed: 11104567]
- Menard S, Fortis S, Castiglioni F, Agresti R, Balsari A. HER2 as a prognostic factor in breast cancer. *Oncology*. 2001; 61(Suppl 2):67–72. [PubMed: 11694790]
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*. 1990; 250(4988):1684–9. [PubMed: 2270482]
- Feilottter HE, Coulon V, McVeigh JL, Boag AH, Dorion-Bonnet F, Duboue B, et al. Analysis of the 10q23 chromosomal region and the PTEN gene in human sporadic breast carcinoma. *Br J Cancer*. 1999; 79(5–6):718–23. [PubMed: 10070859]
- Shimada H, Stram DO, Chatten J, Joshi VV, Hachitanda Y, Brodeur GM, et al. Identification of subsets of neuroblastomas by combined histopathologic and N-myc analysis. *J Natl Cancer Inst*. 1995; 87(19):1470–6. [PubMed: 7674334]

9. Wang S, Gao J, Lei Q, Rozengurt N, Pritchard C, Jiao J, et al. Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell*. 2003; 4(3): 209–21. [PubMed: 14522255]
10. Fujita T, Doihara H, Kawasaki K, Takabatake D, Takahashi H, Washio K, et al. PTEN activity could be a predictive marker of trastuzumab efficacy in the treatment of ErbB2-overexpressing breast cancer. *Br J Cancer*. 2006; 94(2):247–52. [PubMed: 16404430]
11. Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol*. 2002; 20(3):719–26. [PubMed: 11821453]
12. Network TCGA. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490(7418):61–70. [PubMed: 23000897]
13. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012; 486(7403):346–52. [PubMed: 22522925]
14. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98(19):10869–74. [PubMed: 11553815]
15. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JJ, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*. 2010; 12(5):R68. [PubMed: 20813035]
16. Muller WJ, Sinn E, Pattengale PK, Wallace R, Leder P. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. *Cell*. 1988; 54(1):105–15. [PubMed: 2898299]
17. Sinn E, Muller W, Pattengale P, Tepler I, Wallace R, Leder P. Coexpression of MMTV/v-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo. *Cell*. 1987; 49(4):465–75. [PubMed: 3032456]
18. Stewart TA, Pattengale PK, Leder P. Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. *Cell*. 1984; 38(3):627–37. [PubMed: 6488314]
19. Cardiff RD, Anver MR, Gusterson BA, Hennighausen L, Jensen RA, Merino MJ, et al. The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting. *Oncogene*. 2000; 19(8):968–88. [PubMed: 10713680]
20. Andrechek ER, Hardy WR, Siegel PM, Rudnicki MA, Cardiff RD, Muller WJ. Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. *Proc Natl Acad Sci U S A*. 2000; 97(7):3444–9. [PubMed: 10716706]
21. McCarthy A, Savage K, Gabriel A, Naceur C, Reis-Filho JS, Ashworth A. A mouse model of basal-like breast carcinoma with metaplastic elements. *J Pathol*. 2007; 211(4):389–98. [PubMed: 17212342]
22. Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, et al. Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nat Genet*. 1999; 22(1):37–43. [PubMed: 10319859]
23. D'Cruz CM, Gunther EJ, Boxer RB, Hartman JL, Sintasath L, Moody SE, et al. c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nat Med*. 2001; 7(2):235–9. [PubMed: 11175856]
24. Moody SE, Sarkisian CJ, Hahn KT, Gunther EJ, Pickup S, Dugan KD, et al. Conditional activation of Neu in the mammary epithelium of transgenic mice results in reversible pulmonary metastasis. *Cancer Cell*. 2002; 2(6):451–61. [PubMed: 12498714]
25. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc Natl Acad Sci U S A*. 1992; 89(22):10578–82. [PubMed: 1359541]
26. Herschkowitz JJ, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, et al. Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells. *Proc Natl Acad Sci U S A*. 2011

27. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med.* 2011; 17(11):1514–20. [PubMed: 22019887]
28. Hollern DP, Andrechek E. A genomic analysis of mouse models of breast cancer reveals molecular features of mouse models and relationships to human breast cancer. *Breast Cancer Research.* 2014; 16(R59)
29. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* 2007; 8(5):R76. [PubMed: 17493263]
30. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, et al. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol.* 2013; 14(11):R125. [PubMed: 24220145]
31. Silva GO, He X, Parker JS, Gatz ML, Carey LA, Hou JP, et al. Cross-species DNA copy number analyses identifies multiple 1q21-q23 subtype-specific driver genes for breast cancer. *Breast Cancer Res Treat.* 2015; 152(2):347–56. [PubMed: 26109346]
32. Hu G, Chong RA, Yang Q, Wei Y, Blanco MA, Li F, et al. MTDH activation by 8q22 genomic gain promotes chemoresistance and metastasis of poor-prognosis breast cancer. *Cancer Cell.* 2009; 15(1):9–20. [PubMed: 19111877]
33. Ciriello G, Gatz ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell.* 2015; 163(2):506–19. [PubMed: 26451490]
34. Hodgson JG, Malek T, Bornstein S, Hariono S, Ginzinger DG, Muller WJ, et al. Copy number aberrations in mouse breast tumors reveal loci and genes important in tumorigenic receptor tyrosine kinase signaling. *Cancer Res.* 2005; 65(21):9695–704. [PubMed: 16266989]
35. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011; 121(7):2750–67. [PubMed: 21633166]
36. Helmrich A, Stout-Weider K, Hermann K, Schrock E, Heiden T. Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Res.* 2006; 16(10):1222–30. [PubMed: 16954539]
37. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. *Nature.* 2005; 436(7050):518–24. [PubMed: 16049480]
38. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature.* 2006; 439(7074):353–7. [PubMed: 16273092]
39. Gatz ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, et al. A pathway-based classification of human breast cancer. *Proc Natl Acad Sci U S A.* 2010; 107(15):6994–9. [PubMed: 20335537]
40. Ben-David U, Ha G, Khadka P, Jin X, Wong B, Franke L, et al. The landscape of chromosomal aberrations in breast cancer mouse models reveals driver-specific routes to tumorigenesis. *Nature communications.* 2016; 7:12160.
41. Andrechek ER, Cardiff RD, Chang JT, Gatz ML, Acharya CR, Potti A, et al. Genetic heterogeneity of Myc-induced mammary tumors reflecting diverse phenotypes including metastatic potential. *Proc Natl Acad Sci U S A.* 2009; 106(38):16387–92. [PubMed: 19805309]
42. Boxer RB, Jang JW, Sintasath L, Chodosh LA. Lack of sustained regression of c-MYC-induced mammary adenocarcinomas following brief or prolonged MYC inactivation. *Cancer Cell.* 2004; 6(6):577–86. [PubMed: 15607962]
43. Siegel PM, Ryan ED, Cardiff RD, Muller WJ. Elevated expression of activated forms of Neu/ ErbB-2 and ErbB-3 are involved in the induction of mammary tumors in transgenic mice: implications for human breast cancer. *EMBO J.* 1999; 18(8):2149–64. [PubMed: 10205169]
44. Györfy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat.* 2010; 123(3):725–31. [PubMed: 20020197]

45. Andrechek ER. HER2/Neu tumorigenesis and metastasis is regulated by E2F activator transcription factors. *Oncogene*. 2013
46. Fujiwara K, Yuwanita I, Hollern DP, Andrechek ER. Prediction and Genetic Demonstration of a Role for Activator E2Fs in Myc-Induced Tumors. *Cancer Res*. 2011; 71(5):1924–32. [PubMed: 21245101]
47. Yuwanita I, Barnes D, Monterey MD, O'Reilly S, Andrechek ER. Increased metastasis with loss of E2F2 in Myc-driven tumors. *Oncotarget*. 2015
48. Hollern DP, Honeysett J, Cardiff RD, Andrechek ER. The E2F transcription factors regulate tumor development and metastasis in a mouse model of metastatic breast cancer. *Mol Cell Biol*. 2014
49. Andrechek ER, Mori S, Rempel RE, Chang JT, Nevins JR. Patterns of cell signaling pathway activation that characterize mammary development. *Development*. 2008; 135(14):2403–13. [PubMed: 18550711]
50. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A*. 2001; 98(20):11462–7. [PubMed: 11562467]
51. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011; 39(Database issue):D561–8. [PubMed: 21045058]

Figure 1. Identification of copy number alteration through gene expression data from mouse models of breast cancer

(a) The Venn diagram illustrates the consistency of ACE copy number calls with traditional copy number calls of three randomly chose TCGA breast cancer samples. The ACE algorithm was applied to a three TCGA sample and genes that were significantly amplified ($p < .05$, $q < .05$) were compared against the TCGA copy number predictions for the same sample. (b) When applied to the entire TCGA breast cancer dataset with microarray and complete CNV calls ($N=478$) the ACE algorithm is able to identify amplification (top) and deletion (bottom) events with a false discovery rate of 25.4% and 23.3% respectively (left)

and a false negative rate of over 90% (right) (c) The ACE algorithm predicts no copy number alteration in an FVB control mouse. The graph shows neighborhood score, the weighted mean expression value used by the algorithm to predict copy number, at each genomic location (blue line). Significant regions of amplification or deletion are identified when the blue line falls outside of the red bounds ($p < .05$, $q < .05$). (d) Copy number predictions across the genome of a FVB MMTV-Neu tumor reveals notable amplification in chromosome 3 and deletion in chromosome 4 in addition to several smaller amplification and deletion events. (e) When the ACE algorithm is applied to a set of mouse tumors made up of MMTV-Myc tumors (N=12), MMTV-Neu tumors (N=15), MMTV-PyMT tumors (N=26) and DMBA (N=14) and TAG models (N=37), all on the FVB/NJ background, the MMTV-Neu model has significantly ($P < .05$) less amplified or deleted genes compared to the TAG or DMBA models. The MMTV-Neu model was significantly more stable ($P < .05$) and showed on average less copy number changes than all other models.

Figure 2. Landscape of CNA heterogeneity across mouse models of breast cancer

(a) ACE analysis to predict CNA was applied to 600 mouse model tumors arising from 27 major models of breast cancer. The percentage of mice with amplification (red) or deletion (blue) regardless of mouse model at a particular locus across the genome is shown as identified by ACE ($q < .05$). All genes present on the microarray were graphed. Genes that are amplified or deleted in more than 10% of mice are identified. (b) The fraction of mice with amplification or deletion event at individual loci is shown across all chromosomes for individual mouse models. The bar height and color illustrates the percent of samples with an amplification or deletion at the individual gene locus as indicated by the legend.

Figure 3. Conservation of common human CNAs in specific mouse models

(a) To assess the conservation of CNAs in mouse models and human patients unsupervised hierarchical complete linkage clustering of samples across human and mouse tumors were clustered by recurrent CNA events (N=597) that were amplified or deleted in greater than 5% of mouse and human tumors. The dataset used the complete mouse models dataset of 27 mouse models (N=600) and randomly chosen TCGA breast cancer tumors across all five major subtypes of breast cancer (N=559). The clustering revealed three tight clusters composed of human and mouse samples as indicated by the purple, yellow, and green clusters. A highlighted version of this figure is seen (a) while the full version can be found in the supplemental materials. The analysis showed a large fraction of PyMT tumors sorted into each major cluster indicating there are shared copy number alterations between the MMTV-PyMT mouse model and human tumors particularly the Lumina A, Lumina B and

normal like subtypes of breast cancer.(b) Through the use of a Jaccard index we showed within cluster similarity. This reveals a significant decrease of Jaccard Index score when comparing mouse samples from the purple cluster to human samples of the purple, yellow ($P<.05$) and green cluster ($p<.05$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4. Within model CNA heterogeneity associates with tumor histology

(a) For the MMTV-Myc tumor model (n=105), individual tumor samples were clustered by their copy number profile and separated largely into histological subtypes. Histological subtype of each sample is indicated by the color of the bar below the dendrogram as specified by the legend. Vertically the genes (n=4118) are ordered by their chromosomal location from 1 to X. The relationship of the samples is indicated by the dendrogram. The heatmap indicates amplification (red) or deletion (blue) at a particular locus ($p < .05$, $q < .05$). Chromosome 11 amplification was noted in a large fraction of mouse tumor with microacinar histology (Boxed). (b) Mouse tumors with chromosome 11 amplification display a distinct microacinar histological subtype. (c) Human tumors with analogous region (17q25.1) amplified exhibit similar microacinar-like histological patterns. (d) The cbio

oncoprint of the microacinar associated genes across 58 samples. The genes were identified as amplified on mouse chromosome 11 as well as human chromosome region 17q25.1 and total 14 genes in all and identify the core genes associated with microacinar like tumor histology(28 control samples with amplification events and 30 control samples) (e) Across the TCGA breast cancer dataset, patients with a consistent amplification pattern of chromosome 17q23.1 have a microacinar like histological subtype significantly ($P=.01$) more often than those without the amplification event ($N= 28$ for amplified, $N=30$ for non-amplified) indicating a role in this region in defining tumor histology.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5. Role of CNA in tumor metastasis

(a) The schematic outlining the strategy to associate copy number alteration with metastasis score is shown where the metastasis score was correlated with the neighborhood score to assess CNA. Negative correlation (blue) and positive correlation (red) examples are shown. (b) Metastasis associated copy number gains (red) or losses (blue) in the TCGA human breast cancer dataset are shown by human chromosome (B - top). The location of the homologous loci and their amplification status are shown in the mouse model database (B - bottom). Dark black lines indicate example conserved human locations and their associated mouse chromosomal location. Conserved amplification and deletion events between both species are overrepresented when compared to a random set of genes ($P < .0001$). An identified region is conserved between humans chromosome 1 and mouse chromosome 3F indicating a role in tumor metastasis by this region in mouse and human tumors and is more completely explored in panel c. (c) To identify a potential pathway through which metastasis was being mediated we coordinated Ras activity with each amplification or deletion event and looked to identify regions that were associated with metastasis and high ras activity.

This is graphed for mouse chromosome 3F by the negative log of P value for the association of amplification (positive) and deletion (negative) of mouse chromosome 3 where amplification is associated with both metastasis and Ras activity in the 3F region. (d) When tumors are split on the basis of chromosome 3F status those mice, from and MMTV-Myc, MMTV Neu, or MMTV-PyMT background, with amplification (Red) (n=5) are shown to have significantly more metastases than those with a deletion at the same locus (Blue) (n=6) (e). When the region is identified in the kmplot.com dataset this event is shown to be conserved in human Luminal A tumors when the analogous human region is amplified there is significantly (P=.03) higher Ras activity and lower (P<.01) metastasis free survival (f).

Table 1**Composition of dataset**

Table indicating model, number of samples in each model, as well as dataset from which the sample was originally published

Driver	Promoter	Mouse Background	Number of Samples	GSE number
ATX	MMTV	FVB	5	GSE15263
BRCA/p53	Het	C57/B6	12	E-TABM-684
BRCA/p53 (het)	Irradiated	BALBC	7	GSE3165
DMBA	NA	FVB	14	GSE3165
ERBB2	Knock In	FVB	4	GSE37954
IGFIR	MTB	FVB	20	GSE32152
Int3	WAP	FVB	9	GSE3165
LPA	MMTV	FVB	16	GSE15263
Met	MMTV	FVB	52	GSE10450
Myc	MMTV	FVB	12	GSE24594
Myc	MTBTOM	FVB	72	GSE22406
Myc	WAP	FVB	14	GSE3165
Neu	MMTV	FVB	15	GSE42533
Neu	NDL2-5	FVB	21	GSE15904
p53	null	BALBC	17	GSE8863
p53		BALBC	42	GSE3165, GSE27101
p53 (het)	Irradiated	BALBC	7	GSE3165
PyMT	MMTV	FVB	95	GSE3165, GSE30864, GSE20416, GSE10193, GSE23938
PyMT	MMTV	AKXD	16	GSE30864
Ras	MMTV	FVB	10	GSE23938
BLG	Stat3	FVB	16	GSE15119
TAG	C3	FVB	13	GSE16110, GSE25488
TAG	C3	FVB	8	GSE27101
TAG	WAP	BALBC	3	GSE3165
TNP8	WAP	BALBC	35	GSE21444
WNT	MMTV	FVB	39	GSE23938, GSE8828, GSE3165
PYMT	MMTV	FVB	26	Unpublished Data
600				