Breast Cancer
R E S E A R C H

A genomic analysis of mouse models of breast cancer reveals molecular features of mouse models and relationships to human breast cancer

Daniel P Hollern and Eran R Andrechek[*]

Abstract

Introduction: Genomic variability limits the efficacy of breast cancer therapy. To simplify the study of the molecular complexity of breast cancer, researchers have used mouse mammary tumor models. However, the degree to which mouse models model human breast cancer and are reflective of the human heterogeneity has yet to be demonstrated with gene expression studies on a large scale.

Methods: To this end, we have built a database consisting of 1,172 mouse mammary tumor samples from 26 different major oncogenic mouse mammary tumor models.

Results: In this dataset we identified heterogeneity within mouse models and noted a surprising amount of interrelatedness between models, despite differences in the tumor initiating oncogene. Making comparisons between models, we identified differentially expressed genes with alteration correlating with initiating events in each model. Using annotation tools, we identified transcription factors with a high likelihood of activity within these models. Gene signatures predicted activation of major cell signaling pathways in each model, predictions that correlated with previous genetic studies. Finally, we noted relationships between mouse models and human breast cancer at both the level of gene expression and predicted signal pathway activity. Importantly, we identified individual mouse models that recapitulate human breast cancer heterogeneity at the level of gene expression.

Conclusions: This work underscores the importance of fully characterizing mouse tumor biology at molecular, histological and genomic levels before a valid comparison to human breast cancer may be drawn and provides an important bioinformatic resource.

Introduction

Breast cancer is a heterogeneous disease with significant mortality associated with metastatic progression. Classification subdivides human breast cancer into six categories including Luminal A, Luminal B, HER2+, Basal, Claudin-low and normal-like [1]. Recent work suggests additional subclasses exist within each intrinsic subtype including three basal subtypes with striking differences in overall survival [2]. Further, The Cancer Genome Atlas (TCGA) and the Encyclopedia of DNA Elements (ENCODE) projects show remarkable variability in genetic alterations beyond gene expression both across and within subtypes of human breast cancer. Together these genomic analyses demonstrate the complex nature of human breast cancer.

To more readily study mechanisms leading to breast cancer, research has turned to the mouse as a model. Mouse models of breast cancer have employed various methods of initiation, including mouse mammary tumor virus (MMTV) infection, chemical mutagenesis and genetically engineered mice (GEM). This pioneering work identified and tested the role of many oncogenes in breast cancer. With the insertion of MMTV into the genome, numerous key oncogenes were uncovered [3,4]. The later development of MMTV driven transgenics allowed for development of spontaneous models. With the identification of human epithelial growth factor receptor 2 (HER2) amplification in human breast

* Correspondence: andrech1@msu.edu
Department of Physiology, Michigan State University, 2194 Biomedical and Physical Sciences Building, 567 Wilson Rd., East Lansing, MI 48864, USA

cancer [5,6], the observation that MMTV driven expression of the activated rat form of HER2 (NeuNT) resulted in breast cancer reinforced the importance of HER2 as a driving oncogene [7]. More recently, models have been refined to include tissue specific activation resulting in gene amplification, analogous to human HER2+ breast cancer [8], as well as temporal control where transgene expression can be activated or inactivated [9].

Individual mouse models have been used to model aspects of human breast cancer and the selection of the appropriate model to compare to human breast cancer has been directed by phenotype or known genetic events. For instance, the MMTV-PyMT model is widely used to examine metastasis [10] while P53 knockout mammary epithelium transplanted into wild type hosts results in tumors with various genetic mutations [11]. Another aspect is the histological subtype associated with various tumors in GEM models and the metastatic ability can be altered with background [12]. Indeed, similarities between mouse models such as Neu and Wnt as well as their human counterparts have been previously noted [13,14]. Importantly, in both human breast cancer and in many GEM models, there is significant histological heterogeneity [15-17]. These attributes illustrate the importance and utility of mouse models to examine breast cancer.

With the number and variety of GEM models, it is important to consider how accurately these various systems model human breast cancer. Initial studies using intrinsic clustering revealed similarities between mouse models and human breast cancer, albeit in a limited number of samples [18]. Yet, a more detailed characterization of a larger number of p53 null tumors revealed a variety of subtypes with strong similarities to human breast cancer [11], revealing the importance of examining a large number of samples to capture tumor heterogeneity and variability. Further, expanding the number of Myc induced tumors revealed that a subpopulation of Myc induced tumors had similarities to claudin-low human breast cancer [19]. Taken together, recent comparative studies [11,17,19-22] highlighted a clear need for a comprehensive examination of the genomic features of mouse models of breast cancer and their relation to human breast cancer. To this end, we assembled an expansive dataset of mouse models of breast cancer. This dataset reveals the genomic heterogeneity of mouse models and offers a predictive resource for essential cell signaling pathways. Importantly, all comparisons between all models are made available with our report. These data demonstrate the similarities and differences of the various subtypes of mouse models to the key subtypes of human breast cancer and underscore the necessity for an informed choice of the appropriate

mouse model for studying specific types of human breast cancer.

Methods
Combination of datasets
Datasets (GSE10450, GSE11259, GSE13221, GSE13231, GSE13259, GSE13553, GSE13916, GSE14226, GSE14457, GSE14753, GSE15119, GSE15263, GSE15632, GSE15904, GSE16110, GSE17916, GSE18996, GSE20465, GSE20614, GSE21444, GSE22150, GSE22406, GSE23938, GSE24594, GSE25488, GSE27101, GSE30805, GSE30866, GSE3165, GSE31942, GSE32152, GSE34146, GSE34479, GSE6453, GSE6581, GSE6772, GSE7595, GSE8516, GSE8828, GSE8863, GSE9343, GSE9355 GSE37954, GSE2034, GSE2603, GSE4922, GSE6532, AND GSE14020) were downloaded from Gene Expression Omnibus. E-TABM-683 and E-TABM-684 were downloaded from Array Express. For Affymetrix data, Bayesian Factor Regression Methods (BFRM) [23] were used to combine datasets and remove batch effects [24]. Agilent data was merged with Affymetrix data using Chip Comparer [25] and Filemerger [26]. To remove platform effects between Affymetrix and Agilent data and batch effects between individual Agilent studies we used COMBAT [27,28]. Batch effects and batch correction were visualized by principle component analysis in Matlab (for code see Additional file 1).

Data analysis
Unsupervised hierarchical clustering was done using Cluster 3.0 and exported using Java Tree View. The color scheme for the heatmap and sample legends were made using Matlab. Human breast cancer sample intrinsic subtypes were classified according to protocol [1]. Prior to clustering mouse models with human breast cancer, we clustered the human breast tumor samples on their own, to identify genes that would organize the breast tumors according to their intrinsic subtype in the combined dataset. We used these genes to filter the mouse and human combined gene expression dataset for unsupervised hierarchical clustering.

Significance analysis of microarrays [29] was used for fold change analysis. Settings for each comparison can be found in the excel download for each model (Additional files 2, 3). Gene ontology and TRANSFAC predictions were made using GATHER [30]. Gene set enrichment analysis was conducted using Genepattern [31]. The gene-set describing mammary cell-types was derived from [32].

Pathway activation was predicted according to previous studies [2,33]. For mouse samples, specific conditions for each pathway signature can be found in Additional file 4. For human breast tumor samples, pathway activation was predicted using Score Signatures [34] and conditions can be found [2]. Mixture modeling was implemented according to [2].

Results

Database assembly

We assembled a database containing 1,172 samples from mouse mammary tumor models, cell types and normal mammary gland. The major mouse models and descriptions are listed in Table 1. Within a number of these models, variants exist with different alleles, promoters, and genetic backgrounds. In assembling the database, we measured the non-biological variance between gene expression studies and batch correction with principle components analysis (PCA) (Additional file 5A-D). PCA demonstrated that normalization successfully removed artificial variance between datasets (Additional file 5B,D). As a control, we confirmed batch correction utilizing Neu-initiated tumors

spanning the Affymetrix and Agilent platforms from several studies. Prior to normalization (Additional file 5E) PCA demonstrated that Neu tumors varied by platform. After correction, Neu tumors clustered together in PCA, demonstrating that artifactual variance has been removed (Additional file 5F). With platform and batch effects eliminated, we began to explore relationships in the mouse model database.

Gene expression heterogeneity in mouse models

Using unsupervised hierarchical clustering, we examined mouse mammary tumors initiated by various oncogenes. Unsupervised hierarchical clustering generated four major clusters (Figure 1A). We observed remarkable variability

Table 1 List of mouse models in the dataset

Model	Arrays	Promoter	Description	References
Myc	319	MMTV WAP/Dox	Myc mammary tumors of various histological types, expression levels and stability with variable Kras mutations.	[15,19,20,35-38]
Neu	124	MMTV	Induction of adenocarcinomas with pulmonary metastasis.	[11,15,38-44]
PyMT	119	MMTV K6/RCAS MMTV/ RCAS	Rapid induction of luminal-type mammary tumors with pulmonary metastasis.	[38,45-48]
SV40 Large T Antigen	107	C3 WAP	Induction of mammary tumors with similarities to human basal type breast cancer.	[11,38,49-51]
p53	92	Null	Tumors with similarities to human basal type breast cancer.	[11,38,52,53]
CreEtv6/NTRK3	63	WAP	Fusion oncoprotein transforms through activation of AP1.	[54]
MET	52	MMTV	Diverse histologies with similarities to human basal breast cancer.	[16]
BRCA/p53	46	WAP MMTV BLG	CKO of BRCA1 in a p53 null background. Tumors similar to human basal breast cancer.	[38,55]
Wnt	35	MMTV	Induction of mammary tumors with diverse gene expression patterns.	[38,56-58]
IGF-IR	26	MTB	Basal-like mammary tumors. Recurrent tumors resemble human claudin-low.	[59]
LPA	16	MMTV	ER positive, metastatic tumors.	[60]
Stat5	16	BLG	Induction of mammary tumors	NA
Brg1 (+/−)	14	Mutant	Heterogeneous breast cancers.	[61]
DMBA	12	Chemical	Mammary carcinomas with three phenotypes: adenocarcinoma, squamous cell carcinoma, and myoepithelial cell carcinoma.	[62]
Ras	10	MMTV	Induction of mammary tumors with rapid tumor onset.	[38]
Int3	9	WAP	Metastatic tumors.	[63]
RB/p107	7	CKO	Adeno and adenosquamous carcinomas similar to luminal B or basal.	[64]
APC CKO	6	K14-Cre	CKO results in adenocarcinomas with histological and molecular heterogeneity.	[65]
Autotaxin (ATX)	5	MMTV	ER + metastatic mammary tumors.	[60]
BRCA	5	CKO	Tumors similar to human basal type breast cancer.	NA
STAT1	5	Knockout	ERa + PR+, hormone dependent like human ERa + luminal.	[66]
Notch	4	Dox	Induction of mammary adenocarcinomas.	NA
PDK1	2	MMTV	Induction of mammary tumors	[67]
Normal Mammary Gland	47	Not Applicable	Normal mammary gland samples from FVB, BalbC, and CD1 genetic backgrounds.	[16,44,60]

CKO, conditional knockout; Dox, doxycycline inducible MMTV-Rtta system; MMTV, mouse mammary tumor virus.

in gene expression profiles, including within model heterogeneity. For example, Myc initiated tumors span each of the major clusters in the dendrogram. In contrast, some models show uniformity in gene expression from tumor to tumor, including Ras initiated tumors that ordered into a single cluster. Interestingly, there was significant interrelatedness between tumor models initiated with different oncogenes. Annotations for individual tumors revealed that similarities in tumor histology correlated with relationships in gene expression profiles. For example, MMTV-Myc, MMTV-Met and a subset of 7,12-dimethylbenz[a]anthracene (DMBA) induced tumors of the adenosquamous histology shared gene expression profiles. These data reveal mouse models with various levels of heterogeneity and illustrate some of the tumor phenotypes that drive relationships between different mouse models.

To define the characteristics of each cluster, we used Significance Analysis of Microarrays (SAM) to identify differentially regulated genes that define tumors within each cluster (Additional file 6). We interrogated gene lists for gene ontologies (Additional file 6). For instance, Figure 1B shows the gene ontologies for the upregulated genes in the blue cluster in Figure 1A. Ontological categories included genes involved in biological processes and metabolism. To refine these results, tumors from each cluster were examined with Gene Set Enrichment Analysis (GSEA) (Additional file 7). Focusing on tumors in the black cluster, GSEA showed enrichment for gene sets separating mesenchymal cells from luminal cells (Figure 1C, Additional file 8A), including low expression of Zeb1 target genes (Additional file 8B). Gene lists that define mammary stem cells demonstrated that this cluster also had a gene expression profile enriched for mammary stem cell-like features (Additional file 8C,D). In agreement, the majority of epithelial to mesenchymal transition (EMT) like tumors were observed in the black cluster (Figure 1A, Additional file 9). GSEA also demonstrated that tumors from the other clusters had gene expression profiles consistent with luminal cells (Additional file 10). For example, tumors within the blue cluster correlated with gene signatures for luminal progenitor cells and the orange cluster had similarities in gene expression to mature luminal cells. Together, these results define the characteristics of the tumors contained in the major clusters.

Fold change analysis
Given that unique initiating events in the tumor models should cause characteristic responses associated with the tumor initiating event, we used SAM to identify genes significantly altered within each model compared to all other models (Additional file 2). Fold change differences were also calculated between the tumors within a model and normal mammary glands in the corresponding genetic background (Additional file 3). As an example, we determined fold change gene expression differences for Neu initiated tumors (Figure 2A). Collectively, SAM analysis provided a collection of genes that are differentially expressed in each model.

To identify possible transcription factors that could be active in mediating these gene expression changes, we annotated fold change results for each model using TRANSFAC (Additional file 2, 3). For example, for genes regulated by Neu (Figure 2A), we predicted that a significant number of genes had predicted binding sites for the Krox family of transcription factors (Figure 2B). The complete results for the transcription factor binding predictions are included in the additional data for each of the models.

We also annotated fold change differences between each model using gene ontologies (Additional files 2, 3). As an example of the utility of the method, we examined the similarities and differences in gene ontologies in the Neu and TAG models (Figure 2C). Both Neu and TAG tumors featured biological processes, metabolism and nucleic acid-related metabolism as major ontological categories. Key differences included Neu tumors with genes related to transport, ion transport and biosynthesis, categories not found with TAG gene expression changes. TAG tumors had major ontologies representing genes involved in cell cycle, cell organization, cytoskeleton organization and biogenesis, and cell organization and biogenesis. To expand upon gene ontology results we compared each model to all other models and separately to normal mammary gland using GSEA (Additional file 11). This analysis predicted unique features for all models including specific information on metabolism, microenvironment, metastasis and possible pathway activation (Figure 3). For example, TAG tumors had down regulation of genes significantly enriched for the citric acid cycle TCA) (Figure 3A). Wnt tumors were predicted to have upregulation of tumor angiogenesis (Figure 3B). Not surprisingly, polyoma middle T (PyMT) tumors show enrichment for gene sets that predict metastasis (Figure 3C). Finally, GSEA results predicted that p53 mutant tumors may have increased TNF signaling activity (Figure 3D). Together, these results provide a catalogue of possible important features corresponding to the transcriptional outcomes of an initiating oncogene event.

Pathway analysis
To expand the predictive analysis, we utilized a gene signature approach to predict pathway activation across mouse mammary tumors. The pathway prediction relationships between the various models were organized with unsupervised hierarchical clustering (Figure 4). Using this approach, we noted a large degree of heterogeneity within models. Myc tumors showed extensive variation in pathway activation profiles, spanning the spectrum of
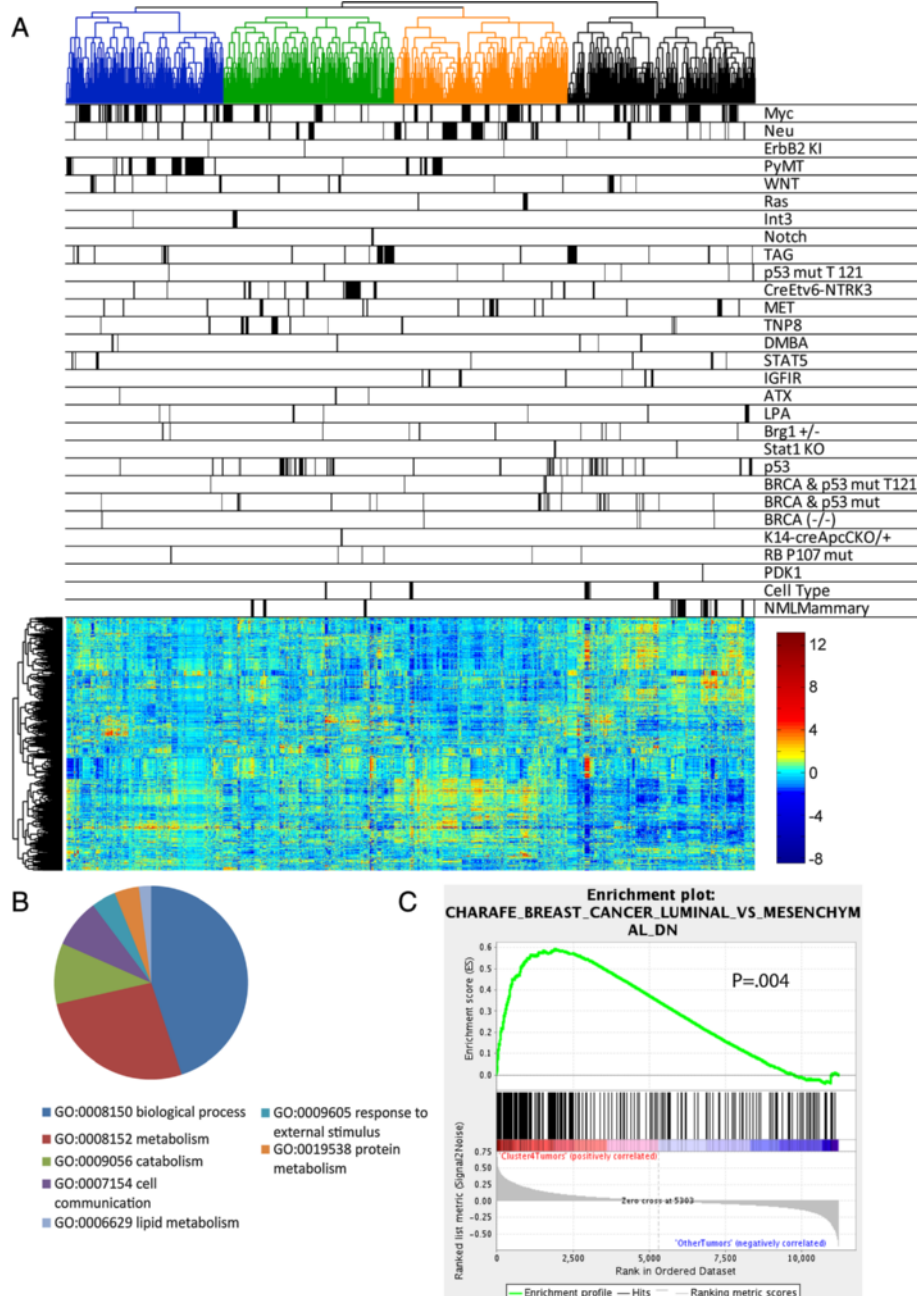
Figure 1 Analysis of relationships between mouse mammary tumor models. (A) The unsupervised hierarchical clustering analysis of gene expression data for mouse mammary tumors, cell types and normal mammary gland is shown. The dendrogram across the top illustrates relationships between samples and is color-coded to itemize the four main clusters. Below the dendrogram, black bars label samples from each corresponding model on the same line. Gene expression values are illustrated with the heatmap, according to the scale shown. The vertical dendrogram beside the heatmap illustrates genes with similar patterns of expression across the samples in the dataset. **(B)** The pie chart illustrates the gene ontologies of the genes that are significantly (q = 0, fdr = 0) over-expressed as identified by SAM in the blue cluster of tumors compared to tumors in other clusters. **(C)** The gene set enrichment plot comparing tumors from cluster 4 (black) to tumors in the other clusters shows significant enrichment for high expression of a gene set that defines mesenchymal breast cancer (P = .004). SAM, significance analysis of microarrays.

clusters. To understand better the heterogeneity and pathway activity within each model, we viewed the pathway predictions on a model-by-model basis (Additional file 12). For example, in PyMT induced tumors, there is a significant difference in predicted pathway activity between tumors from a FVB and AKXD genetic background (Additional file 13). Myc induced tumors with an EMT or squamous histology had distinct predicted pathway activities relative to
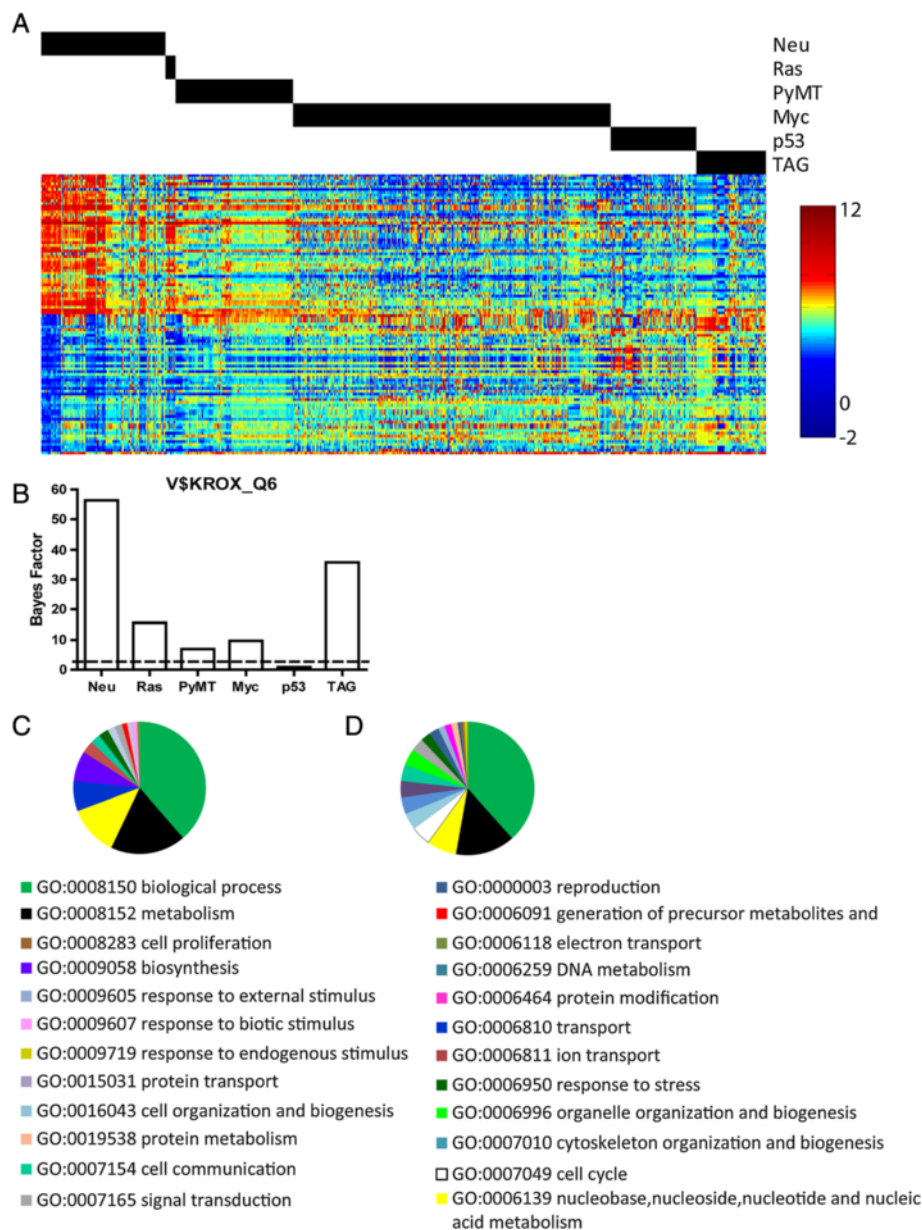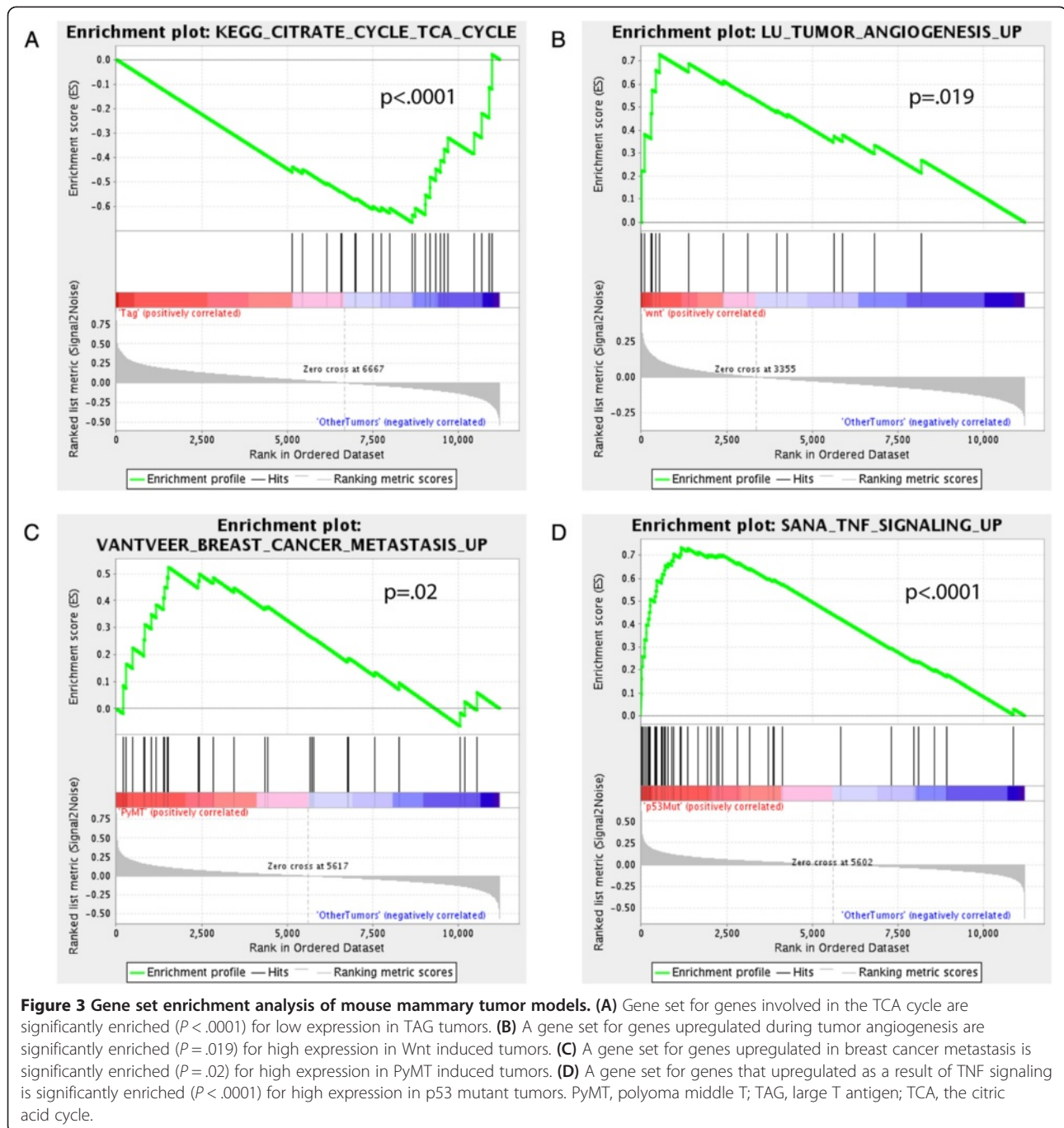
Figure 2 Fold change analysis of Neu induced tumors compared to other tumor models. (A) The expression pattern for the top 50 significantly (q = 0, fdr = 0) upregulated and down regulated genes for Neu-induced tumors as identified by SAM are illustrated with the heatmap. Above the heatmap, black bars denote the model each sample corresponds to. Expression levels are depicted according to the colorbar beside the heatmap. **(B)** The bar graph shows the bayes factor measuring the enrichment of predicted binding sites for the Krox family of transcription factors within upregulated genes from each model. The dotted line indicates a bayes factor of 2.0. **(C)** Gene ontologies for upregulated genes in Neu induced tumors are depicted in the pie chart according to the color-coded categories. **(D)** Gene ontologies for upregulated genes in TAG induced tumors are depicted in the pie chart according to the listed color-coded categories. SAM, significance analysis of microarrays; TAG, large T antigen.

tumors with a papillary or microacinar histology (Additional file 14). In Neu-induced tumors, we observed a major difference in predicted pathway activity between Neu tumors using the MMTV promoter and a Tet-on system to drive oncogene expression (Additional file 15). Taken together, these data demonstrate that tumor type, genetic background, and promoter result in key differences in pathway activity.

To validate and illustrate the utility of pathway activation predictions for developing hypotheses about pathways that function in tumor progression, we identified models with clear pathway activity predictions. Previous genetic

Figure 3 Gene set enrichment analysis of mouse mammary tumor models. (A) Gene set for genes involved in the TCA cycle are significantly enriched ($P < .0001$) for low expression in TAG tumors. **(B)** A gene set for genes upregulated during tumor angiogenesis are significantly enriched ($P = .019$) for high expression in Wnt induced tumors. **(C)** A gene set for genes upregulated in breast cancer metastasis is significantly enriched ($P = .02$) for high expression in PyMT induced tumors. **(D)** A gene set for genes that upregulated as a result of TNF signaling is significantly enriched ($P < .0001$) for high expression in p53 mutant tumors. PyMT, polyoma middle T; TAG, large T antigen; TCA, the citric acid cycle.

studies that correlate with these predictions are noted (Table 2). Demonstrating the validity of the gene signatures, we observe a large degree of agreement between pathways with predicted activity and results from previous investigations.

Comparisons to human breast cancer

With identification of pathways that function in tumor progression in mouse models, it is important to understand whether the given model is reflective of human breast cancer. To this end, we combined datasets for human breast cancer and the mouse mammary tumors in our database, removing both batch and platform effects (Additional file 16). To investigate the relationships between the mouse mammary tumors and human breast tumors, we used unsupervised hierarchical clustering. We identified a large number of mouse mammary tumor models that had similarities in gene expression profiles to human breast cancer (Figure 5). Importantly, Myc and Met induced tumors both recapitulate the
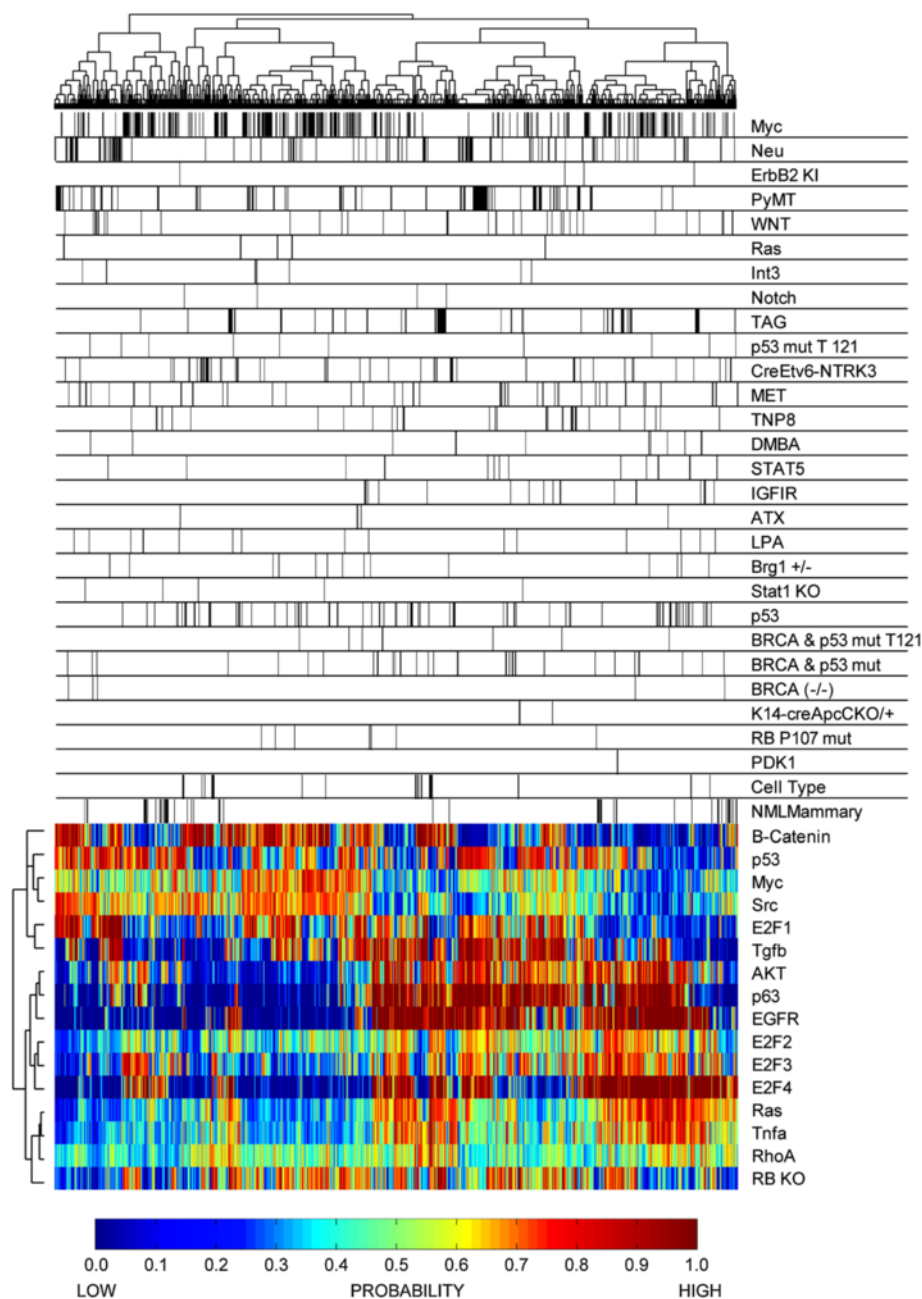
Figure 4 Unsupervised hierarchical clustering of pathway activation predictions in mouse mammary tumors. The dendrogram across the top illustrates the relationship between samples based on predicted pathway activation profiles. Below the dendrogram, the black bars mark tumor samples corresponding to the model listed on the same line. The heatmap illustrates the probability of pathway activation according to the color bar provided below the heatmap. The vertical dendrogram beside the heatmap illustrates pathways with similar predicted activity across the samples in the dataset.

heterogeneity observed in human breast cancer. Using histological annotations, specific relationships between Myc tumor types and human breast cancer subtypes were observed (Additional file 17). For example, Myc tumors with an EMT histology clustered together with human claudin low breast cancer. Extending this to the cluster of tumors predicted to have mesenchymal gene expression features (Figure 1C), we observed that a large majority of these tumors also clustered with claudin low breast cancer. Importantly, further investigation of these tumors matched marker expression for claudin low tumors (Additional file 18A-K). Together these data demonstrated that there are mouse models that share human breast cancer heterogeneity with individual tumor types

Table 2 Validation of pathway predictions

Model	Pathway	Effect	References
APC cKO	B-Catenin	Demonstrated high activation of β-catenin signaling in these tumors.	[65]
APC cKO	Myc	High levels of Myc demonstrated by IHC in these mammary tumors.	[65]
BRCA & P53 mut	EGFR	Using IHC, EGFR was shown to be overexpressed in this mouse model.	[68]
DMBA	Ras	Observation of H-Ras mutations in mammary hyperplastic outgrowths after treatment with DMBA.	[69]
DMBA	EGFR	Using western blot and IHC, EGFR signaling was shown to be active in DMBA induced mammary tumors.	[70]
ETV6-Ntrk3	Src	ETV6-Ntrk3 binds to and activates c-Src, and inhibition of c-Src activation blocks EN transforming activity using mouse engineered mouse embryonic fibroblasts.	[71]
Myc	Ras	Activating mutations in K-Ras found in a subset MMTV-Myc induced tumors with a predicted elevation of Ras signalling.	[15]
Myc	B-Catenin	IHC analysis demonstrates higher expression of B-Catenin in the microacinar histology of Myc driven tumors.	[15]
Myc	E2F1 E2F2 E2F3	E2F loss altered tumor latency and Myc proliferative effects on the mammary gland.	[20]
Neu	Akt	Akt loss effects tumor development in the MMTV-Neu mouse model.	[72]
Neu	B-Catenin	Using a beta-gal reporter, ß-catenin/TCF-dependent transcription was shown to be elevated in MMTV-Neu mouse mammary glands.	[73]
Notch	B-Catenin	Knocking down Notch in a human breast cancer cell line also impacted levels of beta-catenin.	[74]
PyMT	Tgfb	Blockade of TGF-beta inhibits mammary tumor metastasis.	[75]
PyMT	Src	Loss of c-Src greatly reduced the occurrence of mammary tumors in the MMTV-PyMT mouse model.	[76]
Tag	Ras	K-ras amplifications observed in large t-antigen mediated tumorigeneis.	[77]
Tag	E2F2 E2F3 RB KO	Large T Antigen simulates loss of Rb by leading to deregulated acitvation of the E2F transcription factors.	[78]
Wnt	p53	MMTV-Wnt1 mammary tumors with mutant p53 exhibited a superior clinical response compared to tumors with wild-type p53.	[79]

DMBA, 7,12-dimethylbenz[a]anthracene; EGFR, epidermal growth factor receptor; IHC, immunohistochemistry; MMTV, mouse mammary tumor virus; PyMT, polyoma middle T; TGF, transforming growth factor.

that are closely related to subsets of human breast cancer at the level of gene expression.

In addition to comparing mouse mammary tumors and human breast cancer with gene expression, we tested relationships using pathway activation predictions. Using a mixture modeling approach, we clustered human breast cancer into ten different groups based on pathway activation profiles (Figure 6). The pie chart above each heatmap shows the spectrum of the intrinsically annotated samples in each group. No single group was made up of one intrinsic subtype, illustrating the heterogeneity of pathway activation within and between intrinsic subtypes of breast cancer. After groups of human tumors were identified, the probability that an individual mouse mammary tumor belonged to a group of human breast cancer was calculated using the pathway activation profile of the mouse mammary tumor sample. Observing these probabilities with a heatmap, we noted that no single group of human breast cancer was modeled by a single mouse mammary tumor type at the pathway level. Instead, for each group of

human breast cancer, multiple mouse models showed similar predicted pathway activation profiles. Further, these results demonstrated that mouse model relationships to human breast cancer extended beyond the initiating oncogene. For example, mouse tumors initiated by Myc overexpression contained several different tumor types, each modeling a different group of human breast cancer including those groups that have lower predicted Myc activity. Moreover, Neu initiated tumors using an inducible promoter frequently model a single group of human breast cancer (Additional file 19), while other Neu models have diverse pathway activation profiles leading to relationships with several different groups of human breast cancer. These results considered together highlight the similarity and differences between mouse models and human breast cancers.

Discussion

Here we have described the genomic analysis of a dataset composed of publicly available gene expression data for
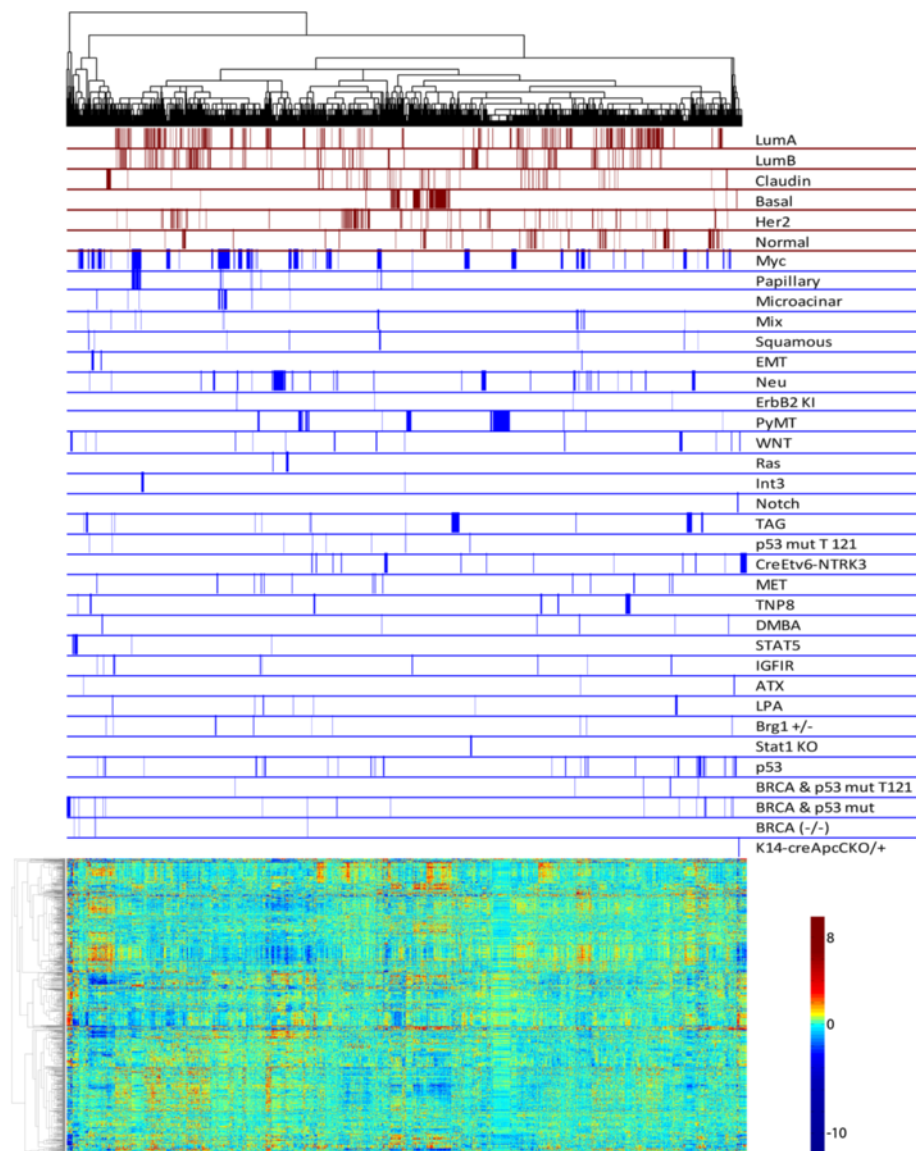
Figure 5 Unsupervised hierarchical clustering of mouse mammary tumor and human breast cancer gene expression data. Across the top, the dendrogram illustrates the relationship between human and mouse tumor samples on the basis of gene expression profiles. The red bars mark the intrinsic subtype of each human tumor sample according the annotation on the same line. The blue bars correspond to the mouse mammary tumor type. Below this, a heatmap shows the gene expression patterns for each sample, with expression values illustrated according to the color bar on the right. The dendrogram beside the heatmap shows the correlation between genes based on expression patterns across the samples in the dataset.

mouse models of breast cancer. These data have been analyzed through a variety of mechanisms to ask how mouse models are distinct, what properties they share and how they reflect human breast cancer. These data indicate that great care should be taken to appropriately choose the mouse model to use and that a genomic and histological characterization of tumors should be completed following experimentation.

In the examination of mouse models in the database, unsupervised hierarchical clustering revealed significant heterogeneity both between models and within models

and was pronounced in tumor models with a large number of samples. Between model differences were fully expected given the unique initiating events causing tumor formation. However, prior studies with relatively few samples for each model did not demonstrate extensive within model heterogeneity [18]. In comparison, we have demonstrated extensive heterogeneity within many models. In part this is due to differences between intrinsic clustering methods [80] and unsupervised hierarchical clustering. However, given that we have noted corresponding differences in fold change, GSEA predictions and pathway signature
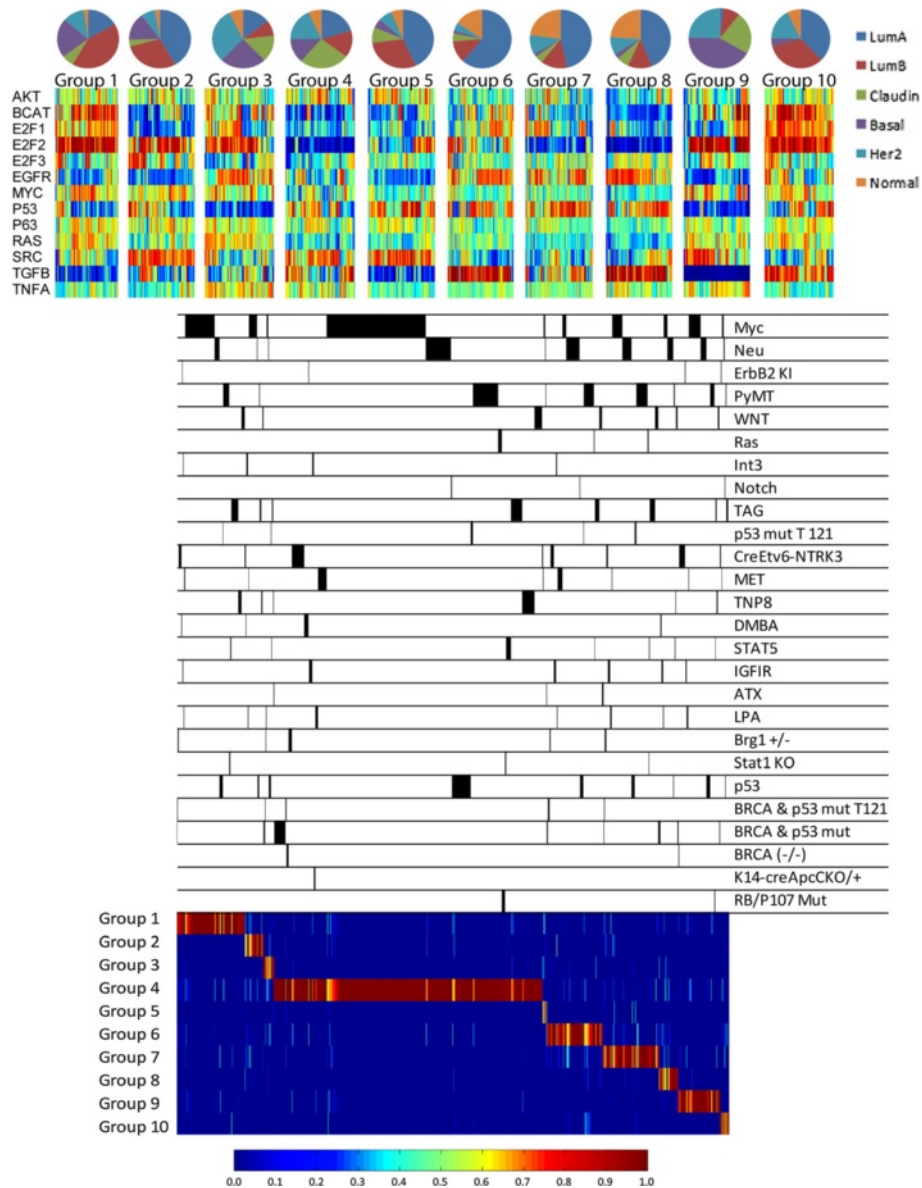
Figure 6 Mixture modeling analysis of human breast cancer pathway heterogeneity and relationships to mouse models of breast cancer. Pie charts above each heatmap illustrate the distribution of the intrinsic subtype of samples in each group, according to the color-coded legend. The heatmap for groups 1 to 10 shows predicted pathway activity with probabilities corresponding to the color bar at the bottom of the figure. Below this, black bars mark the samples corresponding to annotations on the same line. Following the samples down to the heatmap below the black bars, the probability that a mouse model has similar pathway activation profiles is shown for each group. Probabilities for this heatmap are shown according to the color bar at the bottom of the figure.

probabilities, it is likely that this is a reflection of the number of samples used in the analysis. As such, this provides an important caution to characterize a sufficiently large population of tumors to capture heterogeneity in the analysis.

Given that there is typically a predominant histological pattern associated with a given GEM tumor type [81], it is not surprising that there is a predominant genomic pattern. Indeed, we noted for many models that histology is predictive of the genomic subtype. Interestingly, this

histological and genomic interaction is capable of spanning tumor initiating events from different mouse models. Indeed, EMT and spindle-type tumors from diverse models clustered together and were distinct from the non-EMT samples originating in the same model system. Thus, it is also critical for investigators to analyze all tumors from a given model for both histological and genomic patterns.

Mouse models were also investigated individually in comparison to the entire dataset using a variety of methods. This

revealed characteristic gene expression patterns at the fold change level, specific GSEA enrichment effects and key pathway signature differences. In many cases, these results correlated with prior studies. For instance, annotation of fold change results predicted that Neu induced tumors upregulated Krox 20 which is consistent with previous chromatin immunoprecipitation (ChIP) results [82]. When pathway signatures were examined, there were a large number of predictions that could be made for pathways used in specific GEM tumor models. Importantly, while these pathway signatures have previously been validated [2], the model by model pathway predictions shown in Table 2 are highly consistent with previously published tests. For instance, the pathway signatures predicted a high probability of Src activation in PyMT tumors in the FVB background and recent work has demonstrated the necessity for c-Src in PyMT induced tumors [76]. Collectively, for the pathways listed in Table 2, we note agreement between the pathway signature predictions and the reported genetic crosses. Moreover, the pathway signature predictions are also reflective of additional mutations that accumulate in the samples. This was noted in the Myc and TAG induced tumors where the Ras signature was predicted to be elevated, consistent with the large number of Ras activating mutations in these strains [15,77]. Given that numerous published genetic tests are in agreement with the pathway predictions, the remaining cell signaling pathway predictions offer a large number of testable hypotheses. In the future, pathway predictions in the various models should prove to be an important resource for initiating studies into investigating the importance of various signaling pathways in tumor biology.

One of the key aspects of this study was the comparison between mouse models and human breast cancer. These data demonstrated similarities and differences between the two groups and should serve as an important consideration when attempting to extend the comparison of mouse models to human cancer. Taking into account the clustering data, we readily noted that the heterogeneity between human breast cancer samples was present within individual mouse models. Despite capturing the genomic diversity of the samples, we noted several samples with no genomic similarity to human breast cancer, including tumors from strains with other samples that had clear similarity to human breast cancer. This clearly suggests that if conclusions are to be drawn from mouse models of breast cancer, that the mouse samples should be compared and clustered with a variety of human tumors.

In addition to clustering of genomic data, we compared mouse models to human breast cancer through signaling pathway activation predictions. These results showed that for any given group of human breast cancer samples, there was a mouse model with similar pathway activation profiles. Using these results, it is possible to select the mouse model that most closely represents a group of human breast cancer for the signaling pathways of interest. However, it is critical to consider both clustering and pathway activation and to combine these methods to choose the most appropriate model to mimic human breast cancer. For example, to model HER2+ breast cancer and to study the role of HER2 in tumor development, research initially used the MMTV-Neu mice [7]. However, the gene expression data reveals that this strain does not associate with the HER2+ human samples through genomic clustering. However, mixture modeling indicated that a proportion of HER + human cancers did group with the MMTV-Neu samples at the level of pathway activation. This indicates that in some aspects the mouse model is appropriately related to human HER2+ breast cancer. Further, recent reports demonstrate that a strain of mice with conditional activation of Neu under the control of the endogenous promoter which undergo amplification [8] far more closely recapitulate human HER2+ breast cancer [21]. Taken together, these data illustrate the importance of fully characterizing and using all genomic information to select the appropriate model for examination.

Recent reports have described the development of serially transplantable human breast cancer samples that are grown in a murine host with clear genomic similarity to the primary human breast cancer samples [83] and obviously this is an optimal model for specific studies. However, there is clear utility for GEM models, especially with regard to the ability to ask defined genetic questions with regard to key signaling pathways in tumor biology. As such, the prior characterization of mouse and human breast cancer similarities was a critical development [18]. The expanded number of samples and methods of analysis in this report have clearly illustrated additional components of mouse breast cancer biology that require careful consideration. Indeed, the extent of genomic heterogeneity was only appreciated previously for select models [11,15-17], but our work indicates that this is a general characteristic across the majority of breast cancer model systems. As such, this work underscores the requirement to fully characterize mouse tumor biology at histological and genomic levels before a valid comparison to human breast cancer may be drawn. Thus, we have provided the complete files for all of the comparisons made in this manuscript, from fold change between models to GSEA and pathway predictions, with the intent of this being used as a resource to choose and compare mouse models in breast cancer research.

Conclusions

Collectively, our work demonstrates genomic heterogeneity in mouse mammary tumor models. As an additional

outcome of this research, we have provided a large scale predictive resource for each of the mouse models in the database. With heterogeneity driving a variety of relationships between individual mouse mammary tumors and human breast cancer, this work highlights the necessity of fully characterizing mouse tumor biology at molecular, histological and genomic levels before a valid comparison to human breast cancer may be drawn.

Additional files

Additional files 1, 2, 3, 6, 7, 11, 12 are available for download at: https://www.msu.edu/~andrech1/BCR_Supplemental/BCR_Supplemental.html. The results of each analysis are provided as links to zipped folders as described below and are numbered according to their reference in the manuscript. Clicking on a link will begin the download of the zipped material.

Additional file 1: PCA code for Matlab.

Additional file 2: Fold change values organized by mouse tumor model type comparing the mouse model to all other types of mouse models.

Additional file 3: Fold change values organized by mouse tumor model type comparing the mouse model to normal mammary gland.

Additional file 4: Table S1. Gene signature settings for pathway predictions. Specific settings are shown for each gene signature of pathway activation.

Additional file 5: Figure S1. Removal of Batch Effects from Affymetrix Datasets. (A) Affymetrix datasets color coded according to the study of origin in a principle components analysis plot prior to BFRM batch effect correction. (B) Affymetrix datasets color coded according to the study of origin in a principle components analysis plot after BFRM batch effect correction. (C) Affymetrix datasets are color coded together in blue after BFRM batch effect correction. The various Agilent gene expression datasets are color-coded and plotted along with Affymetrix data on the three principle components to illustrate platform and batch variance. (D) Agilent and Affymetrix color-coded data plotted after COMBAT removed batch and platform technical variance. (E) Neu-induced tumors are color coded in blue and all other tumors are in green, illustrating variance between similar tumor types on the basis of platform and batch artifacts. (F) Neu-induced tumors are color coded in blue and all other tumors are in green illustrating mediation of batch and platform effects.

Additional file 6: Fold change for genes in clusters 1-4 from the manuscript and the gene ontology associated with each cluster.

Additional file 7: GSEA for clusters 1-4, each of the runs (C2, C3, and so on) is in a separate folder.

Additional file 8: Figure S2. Gene set enrichment analysis for mouse mammary tumors in the black color-coded cluster. (A) A gene set for down regulated genes in mesenchymal breast cancer is significantly enriched (P <.0001) and down regulated in the black cluster (cluster4) of tumors. (B) A gene set for Zeb1 target genes is significantly enriched (P = .005) for low expression for the tumors in the black cluster. (C) A gene set for genes highly expressed in mammary stem cells is significantly enriched (P = .016) and upregulated in tumors from cluster 4 (black). (D) A gene set for genes that are down regulated in mammary stem cells is significantly enriched (P <.0001) and also down regulated in the cluster 4 (black) tumors.

Additional file 9: Figure S3. Tumors that were classified in mesenchymal histology cluster into the black cluster. Highlighting prior histological annotations for mesenchymal or EMT-like tumors across the

Myc, IGF-IR, DMBA, and p53 mutant models show that a large majority of these tumors cluster together in the black cluster.

Additional file 10: Figure S4. Gene set enrichment analysis for mammary cell types across major clusters of mouse mammary tumors. GSEA for tumors in blue cluster compared to all other clusters show significant enrichment for a mammary luminal progenitor cell gene expression signature (P = .006). Similarly, tumors from the green cluster associate with a mixture of luminal cell gene expression features, while tumors in the orange cluster are significantly enriched for gene expression features of mature luminal cells (P = .04). Lastly, tumors in the black cluster are significantly enriched for gene expression features of mammary stem cells (P = .01).

Additional file 11: GSEA for mouse models compared to all other models or to mammary gland development. Listed by model.

Additional file 12: PDFs of pathway predictions for each mouse model of breast cancer, folders exist for each mouse modelx.

Additional file 13: Figure S5. Unsupervised hierarchical clustering of pathway probabilities for PyMT induced tumors. The dendrogram across the top illustrates the relationship between PyMT tumor types on the basis of pathway activation profiles. Below the dendrogram black bars correspond to sample details on the same line, annotating the genetic background and sample type for each sample. The heatmap shows the predicted pathway activity according to the probabilities listed on the color bar below the heatmap. Directly beside the heatmap, a vertical dendrogram illustrates the degree of correlation between pathways across the samples.

Additional file 14: Figure S6. Unsupervised hierarchical clustering of pathway probabilities for Myc induced tumors. The dendrogram across the top illustrates the relationship between Myc tumor types on the basis of pathway activation profiles. Below the dendrogram black bars correspond to sample details on the same line, annotating the tumor histology (if known), specific form of Myc expression, recurrence status, and additional modifications. The heatmap shows the predicted pathway activity according to the probabilities listed on the color bar below the heatmap. Directly beside the heatmap, a vertical dendrogram illustrates the degree of correlation between pathways across the samples.

Additional file 15: Figure S7. Unsupervised hierarchical clustering of pathway probabilities for Neu induced tumors. The dendrogram across the top illustrates the relationship between Neu tumor types on the basis of pathway activation profiles. Below the dendrogram black bars correspond to sample details on the same line, annotating the specific form of Neu, and additional modifications. The heatmap shows the predicted pathway activity according to the probabilities listed on the color bar below the heatmap. Directly beside the heatmap, a vertical dendrogram illustrates the degree of correlation between pathways across the samples.

Additional file 16: Figure S8. Removal of batch effects between mouse and human breast cancer datasets. (A) Mouse (blue) and human (green) Affymetrix data gene expression variance plotted onto three principle components prior to BFRM. (B) Mouse (blue) and human (green) Affymetrix data gene expression variance plotted onto three principle components after BFRM. (C) Human(green) and mouse (blue) Affymetrix data after BFRM correction put together with mouse Agilent data (red) prior to COMBAT. (D) Human(green) and mouse (blue) Affymetrix data after BFRM correction put together with mouse Agilent data (red) after COMBAT artifact correction.

Additional file 17: Figure S9. Unsupervised hierarchical clustering of Myc mouse mammary tumors and human breast cancer gene expression data. Across the top, the dendrogram illustrates the relationship between human and mouse tumor samples on the basis of gene expression profiles. The red bars mark the intrinsic subtype of each human tumor sample according to the annotation on the same line. The blue bars correspond to the Myc mouse mammary tumor type. Below this, a heatmap shows the gene expression patterns for each sample, with expression values illustrated according to the color bar on the right. The dendrogram beside the heatmap shows the correlation between genes based on expression patterns across the samples in the dataset.

Additional file 18: Figure S10. Claudin low marker expression in the black cluster mouse mammary tumors. Claudin low marker expression comparisons for cluster 4 (black) tumors compared to tumors in all other clusters as defined by Figure 1A. (A-C) Cell adhesion markers that have low expression in claudin low human tumors are also down regulated in cluster 4 (black tumors), $P < .0001$. (D-E) Genes that are involved with the immune system that are found to be highly expressed in claudin low human tumors are highly expressed in mouse cluster 4 tumors (black), $P < .01$ for CD79B and $P < .0001$ for VAV1. (F) Chemokine (C-X-C motif) ligand 12, involved in cell communication and previously shown to be highly expressed in claudin low tumors, is upregulated in cluster 4 (black) mouse mammary tumors, $P < .0001$. (G) Fibroblast growth factor 7, an extracellular matrix related factor and previously shown to be highly expressed in claudin low tumors, is upregulated in cluster 4 (black) mouse mammary tumors, $P < .0001$. (H-J) Cell migration markers previously shown to be highly expressed in human claudin low tumors are upregulated in mouse cluster 4(black) tumors, $P < .02$ for moesin and $P < .0001$ for integrin α5. (K) Angiogenesis marker, VEGFC, was previously shown to be upregulated in human claudin low tumors and is highly expressed in mouse cluster 4(black) tumors.

Additional file 19: Figure S11. Mixture modeling highlighting pathway relationships between human breast cancer and specific models of Neu mediated tumorigenesis. Pie charts above each heatmap illustrate the distribution of the intrinsic subtype of samples in each group, according to the color-coded legend. The heatmap for groups 1 to 10 show predicted pathway activity with probabilities corresponding to the color bar at the bottom of the figure. Below this, blue bars mark the samples corresponding to annotations on the same line. Following the samples down to the heatmap below the blue bars, the probability that a specific type of Neu model has similar pathway activation profiles is shown for each group. Probabilities for this heatmap are shown according to the color bar at the bottom of the figure.

Abbreviations

DMBA: 7,12-dimethylbenz[a]anthracene; EMT: epithelial to mesenchymal transition; ENCODE: Encyclopedia of DNA Elements; GATHER: gene annotation tool to help explain relationships; GEM: genetically engineered mice; GSEA: gene set enrichment analysis; HER2: human epidermal growth factor receptor 2; MMTV: mouse mammary tumor virus; PCA: principle components analysis; PyMT: polyoma middle T.; SAM: significance analysis of microarrays; TAG: large T antigen; TCA: the citric acid cycle; TCGA: The Cancer Genome Atlas; TNF: tumor necrosis factor; TRANSFAC: transcription factor database.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DH and EA conceived of the study, and participated in its design and coordination and helped to draft the manuscript. DH performed the experiments in this study. DH and EA interpreted the data. Both authors read and approved the final manuscript.

References

1. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM: **Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer.** *Breast Cancer Res* 2010, **12**:R68.
2. Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, Datto MB, Kelley M, Mathey-Prevot B, Potti A, Nevins JR: **A pathway-based classification of human breast cancer.** *Proc Natl Acad Sci U S A* 2010, **107**:6994–6999.
3. Callahan R, Smith GH: **MMTV-induced mammary tumorigenesis: gene discovery, progression to malignancy and cellular pathways.** *Oncogene* 2000, **19**:992–1001.
4. Nusse R, Varmus HE: **Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome.** *Cell* 1982, **31**:99–109.
5. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL: **Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.** *Science* 1987, **235**:177–182.
6. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, Press MF: **Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.** *Science* 1989, **244**:707–712.
7. 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**:105–115.
8. 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**:3444–3449.
9. D'Cruz CM, Gunther EJ, Boxer RB, Hartman JL, Sintasath L, Moody SE, Cox JD, Ha SI, Belka GK, Golant A, Cardiff RD, Chodosh LA: **c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations.** *Nat Med* 2001, **7**:235–239.
10. Fluck MM, Schaffhausen BS: **Lessons in signaling and tumorigenesis from polyomavirus middle T antigen.** *Microbiol Mol Biol Rev* 2009, **73**:542–563. Table of Contents.
11. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, Knezevic J, Greene SB, Darr D, Troester MA, Hilsenbeck SG, Medina D, Perou CM, Rosen JM: **Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells.** *Proc Natl Acad Sci U S A* 2011, **109**:2778–2783.
12. Hunter KW, Broman KW, Voyer TL, Lukes L, Cozma D, Debies MT, Rouse J, Welch DR: **Predisposition to efficient mammary tumor metastatic progression is linked to the breast cancer metastasis suppressor gene Brms1.** *Cancer Res* 2001, **61**:8866–8872.
13. Andrechek ER, Laing MA, Girgis-Gabardo AA, Siegel PM, Cardiff RD, Muller WJ: **Gene expression profiling of neu-induced mammary tumors from transgenic mice reveals genetic and morphological similarities to ErbB2-expressing human breast cancers.** *Cancer Res* 2003, **63**:4920–4926.
14. Rosner A, Miyoshi K, Landesman-Bollag E, Xu X, Seldin DC, Moser AR, MacLeod CL, Shyamala G, Gillgrass AE, Cardiff RD: **Pathway pathology: histological differences between ErbB/Ras and Wnt pathway transgenic mammary tumors.** *Am J Pathol* 2002, **161**:1087–1097.
15. Andrechek ER, Cardiff RD, Chang JT, Gatza ML, Acharya CR, Potti A, Nevins JR: **Genetic heterogeneity of Myc-induced mammary tumors reflecting diverse phenotypes including metastatic potential.** *Proc Natl Acad Sci U S A* 2009, **106**:16387–16392.
16. Ponzo MG, Lesurf R, Petkiewicz S, O'Malley FP, Pinnaduwage D, Andrulis IL, Bull SB, Chughtai N, Zuo D, Souleimanova M, Germain D, Omeroglu A, Cardiff RD, Hallett M, Park M: **Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer.** *Proc Natl Acad Sci U S A* 2009, **106**:12903–12908.
17. Knight JF, Lesurf R, Zhao H, Pinnaduwage D, Davis RR, Saleh SM, Zuo D, Naujokas MA, Chughtai N, Herschkowitz JI, Prat A, Mulligan AM, Muller WJ, Cardiff RD, Gregg JP, Andrulis IL, Hallett MT, Park M: **Met synergizes with p53 loss to induce mammary tumors that possess features of claudin-low breast cancer.** *Proc Natl Acad Sci U S A* 2013, **110**:E1301–E1310.
18. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors.** *Genome Biol* 2007, **8**:R76.
19. Hollern DP, Yuwanita I, Andrechek ER: **A mouse model with T58A mutations in Myc reduces the dependence on KRas mutations and has similarities to claudin-low human breast cancer.** *Oncogene* 2012, **32**:1296–1304.

20. 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**:1924–1932.

21. Schade B, Lesurf R, Sanguin-Gendreau V, Bui T, Deblois G, O'Toole SA, Millar EK, Zardawi SJ, Lopez-Knowles E, Sutherland RL, Giguère V, Kahn M, Hallett M, Muller WJ: **beta-catenin signaling is a critical event in ErbB2-mediated mammary tumor progression.** *Cancer Res* 2013, **73**:4474–4487.

22. Dourdin N, Schade B, Lesurf R, Hallett M, Munn RJ, Cardiff RD, Muller WJ: **Phosphatase and tensin homologue deleted on chromosome 10 deficiency accelerates tumor induction in a mouse model of ErbB-2 mammary tumorigenesis.** *Cancer Res* 2008, **68**:2122–2131.

23. Carvalho CM, Chang J, Lucas JE, Nevins JR, Wang Q, West M: **High-dimensional sparse factor modeling: applications in gene expression genomics.** *J Am Stat Assoc* 2008, **103**:1438–1456.

24. http://www.stat.duke.edu/research/software/west/bfrm/download.html.

25. http://chipcomparer.genome.duke.edu/.

26. http://filemerger.genome.duke.edu/.

27. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods.** *Biostatistics* 2007, **8**:118–127.

28. http://www.bu.edu/jlab/wp-assets/ComBat/Download.html.

29. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98**:5116–5121.

30. http://gather.genome.duke.edu/.

31. http://genepattern.broadinstitute.org/gp/pages/login.jsf.

32. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK, Visvader JE: **Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways.** *Breast Cancer Res* 2010, **12**:R21.

33. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA Jr, Marks JR, Dressman HK, West M, Nevins JR: **Oncogenic pathway signatures in human cancers as a guide to targeted therapies.** *Nature* 2006, **439**:353–357.

34. https://genepattern.genome.duke.edu/gp/pages/login.jsf.

35. Wertheim GB, Yang TW, Pan TC, Ramne A, Liu Z, Gardner HP, Dugan KD, Kristel P, Kreike B, van de Vijver MJ, Cardiff RD, Reynolds C, Chodosh LA: **The Snf1-related kinase, Hunk, is essential for mammary tumor metastasis.** *Proc Natl Acad Sci U S A* 2009, **106**:15855–15860.

36. Sandgren EP, Schroeder JA, Qui TH, Palmiter RD, Brinster RL, Lee DC: **Inhibition of mammary gland involution is associated with transforming growth factor alpha but not c-myc-induced tumorigenesis in transgenic mice.** *Cancer Res* 1995, **55**:3915–3927.

37. Leung JY, Andrechek ER, Cardiff RD, Nevins JR: **Heterogeneity in MYC-induced mammary tumors contributes to escape from oncogene dependence.** *Oncogene* 2012, **31**:2545–2554.

38. Zhu M, Yi M, Kim CH, Deng C, Li Y, Medina D, Stephens RM, Green JE: **Integrated miRNA and mRNA expression profiling of mouse mammary tumor models identifies miRNA signatures associated with mammary tumor lineage.** *Genome Biol* 2011, **12**:R77.

39. Ranger JJ, Levy DE, Shahalizadeh S, Hallett M, Muller WJ: **Identification of a Stat3-dependent transcription regulatory network involved in metastatic progression.** *Cancer Res* 2009, **69**:6823–6830.

40. 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**:10578–10582.

41. Schade B, Rao T, Dourdin N, Lesurf R, Hallett M, Cardiff RD, Muller WJ: **PTEN deficiency in a luminal ErbB-2 mouse model results in dramatic acceleration of mammary tumorigenesis and metastasis.** *J Biol Chem* 2009, **284**:19018–19026.

42. Schade B, Lam SH, Cernea D, Sanguin-Gendreau V, Cardiff RD, Jung BL, Hallett M, Muller WJ: **Distinct ErbB-2 coupled signaling pathways promote mammary tumors with unique pathologic and transcriptional profiles.** *Cancer Res* 2007, **67**:7579–7588.

43. Ursini-Siegel J, Cory S, Zuo D, Hardy WR, Rexhepaj E, Lam S, Schade B, Jirstrom K, Bjur E, Piccirillo CA, Denardo D, Coussens LM, Brennan DJ, Gallagher WM, Park M, Pawson T, Hallett M, Muller WJ: **Receptor tyrosine kinase signaling favors a protumorigenic state in breast cancer cells by inhibiting the adaptive immune response.** *Cancer Res* 2010, **70**:7776–7787.

44. Schoenherr RM, Kelly-Spratt KS, Lin C, Whiteaker JR, Liu T, Holzman T, Coleman I, Feng LC, Lorentzen TD, Krasnoselsky AL, Wang P, Liu Y, Gurley KE, Amon LM, Schepmoes AA, Moore RJ, Camp DG 2nd, Chodosh LA, Smith RD, Nelson PS, McIntosh MW, Kemp CJ, Paulovich AG: **Proteome and transcriptome profiles of a Her2/Neu-driven mouse model of breast cancer.** *Proteomics Clin Appl* 2011, **5**:179–188.

45. Guy CT, Cardiff RD, Muller WJ: **Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease.** *Mol Cell Biol* 1992, **12**:954–961.

46. Hu Y, Wu G, Rusch M, Lukes L, Buetow KH, Zhang J, Hunter KW: **Integrated cross-species transcriptional network analysis of metastatic susceptibility.** *Proc Natl Acad Sci U S A* 2012, **109**:3184–3189.

47. Bu W, Chen J, Morrison GD, Huang S, Creighton CJ, Huang J, Chamness GC, Hilsenbeck SG, Roop DR, Leavitt AD, Li Y: **Keratin 6a marks mammary bipotential progenitor cells that can give rise to a unique tumor model resembling human normal-like breast cancer.** *Oncogene* 2011, **30**:4399–4409.

48. Flowers M, Schroeder JA, Borowsky AD, Besselsen DG, Thomson CA, Pandey R, Thompson PA: **Pilot study on the effects of dietary conjugated linoleic acid on tumorigenesis and gene expression in PyMT transgenic mice.** *Carcinogenesis* 2010, **31**:1642–1649.

49. Klein A, Wessel R, Graessmann M, Jurgens M, Petersen I, Schmutzler R, Niederacher D, Arnold N, Meindl A, Scherneck S, Seitz S, Graessmann A: **Comparison of gene expression data from human and mouse breast cancers: identification of a conserved breast tumor gene set.** *Int J Cancer* 2007, **121**:683–688.

50. Maroulakou IG, Anver M, Garrett L, Green JE: **Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene.** *Proc Natl Acad Sci U S A* 1994, **91**:11236–11240.

51. Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W: **Identification of early molecular markers for breast cancer.** *Mol Cancer* 2011, **10**:15.

52. Zhang M, Behbod F, Atkinson RL, Landis MD, Kittrell F, Edwards D, Medina D, Tsimelzon A, Hilsenbeck S, Green JE, Michalowska AM, Rosen JM: **Identification of tumor-initiating cells in a p53-null mouse model of breast cancer.** *Cancer Res* 2008, **68**:4674–4682.

53. Backlund MG, Trasti SL, Backlund DC, Cressman VL, Godfrey V, Koller BH: **Impact of ionizing radiation and genetic background on mammary tumorigenesis in p53-deficient mice.** *Cancer Res* 2001, **61**:6577–6582.

54. Li Z, Tognon CE, Godinho FJ, Yasaitis L, Hock H, Herschkowitz JI, Lannon CL, Cho E, Kim SJ, Bronson RT, Perou CM, Sorensen PH, Orkin SH: **ETV6-NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex.** *Cancer Cell* 2007, **12**:542–558.

55. Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, Hennighausen L, Wynshaw-Boris A, Deng CX: **Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation.** *Nat Genet* 1999, **22**:37–43.

56. Cho RW, Wang X, Diehn M, Shedden K, Chen GY, Sherlock G, Gurney A, Lewicki J, Clarke MF: **Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors.** *Stem Cells* 2008, **26**:364–371.

57. Tsukamoto AS, Grosschedl R, Guzman RC, Parslow T, Varmus HE: **Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice.** *Cell* 1988, **55**:619–625.

58. Pond AC, Herschkowitz JI, Schwertfeger KL, Welm B, Zhang Y, York B, Cardiff RD, Hilsenbeck S, Perou CM, Creighton CJ, Lloyd RE, Rosen JM: **Fibroblast growth factor receptor signaling dramatically accelerates tumorigenesis and enhances oncoprotein translation in the mouse mammary tumor virus-Wnt-1 mouse model of breast cancer.** *Cancer Res* 2010, **70**:4868–4879.

59. Franks SE, Campbell CI, Barnett EF, Siwicky MD, Livingstone J, Cory S, Moorehead RA: **Transgenic IGF-IR overexpression induces mammary tumors with basal-like characteristics, whereas IGF-IR-independent mammary tumors express a claudin-low gene signature.** *Oncogene* 2011, **31**:3298–3309.

60. Liu S, Umezu-Goto M, Murph M, Lu Y, Liu W, Zhang F, Yu S, Stephens LC, Cui X, Murrow G, Coombes K, Muller W, Hung MC, Perou CM, Lee AV, Fang X, Mills GB: **Expression of autotaxin and lysophosphatidic acid receptors increases mammary tumorigenesis, invasion, and metastases.** *Cancer Cell* 2009, **15**:539–550.

61. Bultman SJ, Herschkowitz JI, Godfrey V, Gebuhr TC, Yaniv M, Perou CM, Magnuson T: **Characterization of mammary tumors from Brg1 heterozygous mice.** *Oncogene* 2008, **27**:460–468.

62. Yin Y, Bai R, Russell RG, Beildeck ME, Xie Z, Kopelovich L, Glazer RI: Characterization of medroxyprogesterone and DMBA-induced multilineage mammary tumors by gene expression profiling. *Mol Carcinog* 2005, **44**:42–50.

63. Gallahan D, Jhappan C, Robinson G, Hennighausen L, Sharp R, Kordon E, Callahan R, Merlino G, Smith GH: Expression of a truncated Int3 gene in developing secretory mammary epithelium specifically retards lobular differentiation resulting in tumorigenesis. *Cancer Res* 1996, **56**:1775–1785.

64. Jiang Z, Deng T, Jones R, Li H, Herschkowitz JI, Liu JC, Weigman VJ, Tsao MS, Lane TF, Perou CM, Zacksenhaus E: Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status. *J Clin Invest* 2010, **120**:3296–3309.

65. Kuraguchi M, Ohene-Baah NY, Sonkin D, Bronson RT, Kucherlapati R: Genetic mechanisms in Apc-mediated mammary tumorigenesis. *PLoS Genet* 2009, **5**:e1000367.

66. Chan SR, Vermi W, Luo J, Lucini L, Rickert C, Fowler AM, Lonardi S, Arthur C, Young LJ, Levy DE, Welch MJ, Cardiff RD, Schreiber RD: STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas. *Breast Cancer Res* 2012, **14**:R16.

67. Pollock CB, Yin Y, Yuan H, Zeng X, King S, Li X, Kopelovich L, Albanese C, Glazer RI: PPARdelta activation acts cooperatively with 3-phosphoinositide-dependent protein kinase-1 to enhance mammary tumorigenesis. *PLoS One* 2011, **6**:e16215.

68. 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**:389–398.

69. Kumar R, Medina D, Sukumar S: Activation of H-ras oncogenes in preneoplastic mouse mammary tissues. *Oncogene* 1990, **5**:1271–1277.

70. Kim S, Roopra A, Alexander CM: A phenotypic mouse model of basaloid breast tumors. *PLoS One* 2012, **7**:e30979.

71. Jin W, Yun C, Hobbie A, Martin MJ, Sorensen PH, Kim SJ: Cellular transformation and activation of the phosphoinositide-3-kinase-Akt cascade by the ETV6-NTRK3 chimeric tyrosine kinase requires c-Src. *Cancer Res* 2007, **67**:3192–3200.

72. Maroulakou IG, Oemler W, Naber SP, Tsichlis PN: Akt1 ablation inhibits, whereas Akt2 ablation accelerates, the development of mammary adenocarcinomas in mouse mammary tumor virus (MMTV)-ErbB2/neu and MMTV-polyoma middle T transgenic mice. *Cancer Res* 2007, **67**:167–177.

73. Khalil S, Tan GA, Giri DD, Zhou XK, Howe LR: Activation status of Wnt/ss-catenin signaling in normal and neoplastic breast tissues: relationship to HER2/neu expression in human and mouse. *PLoS One* 2012, **7**:e33421.

74. Wang J, Fu L, Gu F, Ma Y: Notch1 is involved in migration and invasion of human breast cancer cells. *Oncol Rep* 2011, **26**:1295–1303.

75. Muraoka RS, Dumont N, Ritter CA, Dugger TC, Brantley DM, Chen J, Easterly E, Roebuck LR, Ryan S, Gotwals PJ, Koteliansky V, Arteaga CL: Blockade of TGF-beta inhibits mammary tumor cell viability, migration, and metastases. *J Clin Invest* 2002, **109**:1551–1559.

76. Guy CT, Muthuswamy SK, Cardiff RD, Soriano P, Muller WJ: Activation of the c-Src tyrosine kinase is required for the induction of mammary tumors in transgenic mice. *Genes Dev* 1994, **8**:23–32.

77. Liu ML, Von Lintig FC, Liyanage M, Shibata MA, Jorcyk CL, Ried T, Boss GR, Green JE: Amplification of Ki-ras and elevation of MAP kinase activity during mammary tumor progression in C3(1)/SV40 Tag transgenic mice. *Oncogene* 1998, **17**:2403–2411.

78. Taneja P, Zhu S, Maglic D, Fry EA, Kendig RD, Inoue K: Transgenic and knockout mice models to reveal the functions of tumor suppressor genes. *Clin Med Insights Oncol* 2011, **5**:235–257.

79. Jackson JG, Pant V, Li Q, Chang LL, Quintas-Cardama A, Garza D, Tavana O, Yang P, Manshouri T, Li Y, El-Naggar AK, Lozano G: p53-mediated senescence impairs the apoptotic response to chemotherapy and clinical outcome in breast cancer. *Cancer Cell* 2012, **21**:793–806.

80. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D: Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 1999, **96**:9212–9217.

81. Cardiff RD, Anver MR, Gusterson BA, Hennighausen L, Jensen RA, Merino MJ, Rehm S, Russo J, Tavassoli FA, Wakefield LM, Ward JM, Green JE: The mammary pathology of genetically engineered mice: the consensus report and recommendations from the Annapolis meeting. *Oncogene* 2000, **19**:968–988.

82. Dillon RL, Brown ST, Ling C, Shioda T, Muller WJ: An EGR2/CITED1 transcription factor complex and the 14-3-3sigma tumor suppressor are involved in regulating ErbB2 expression in a transgenic-mouse model of human breast cancer. *Mol Cell Biol* 2007, **27**:8648–8657.

83. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, Factor R, Matsen C, Milash BA, Nelson E, Neumayer L, Randall RL, Stijleman IJ, Welm BE, Welm AL: Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med* 2011, **17**:1514–1520.