

GENETICALLY ENGINEERED MOUSE MODELS OF BREAST CARCINOMA:
A TRANSLATIONAL RESOURCE FOR HIGHLIGHTING HUMAN BREAST SUBTYPE
ETIOLOGY AND DEVELOPING PERSONALIZED THERAPEUTIC APPROACHES

Adam David Pfefferle

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department
of Pathology and Laboratory Medicine

Chapel Hill
2015

Approved by:

Charles M. Perou

H. Shelton Earp III

William K. Kaufmann

Joel S. Parker

Norman E. Sharpless

Bernard E. Weissman

© 2015
Adam David Pfefferle
ALL RIGHTS RESERVED

ABSTRACT

Adam David Pfefferle: Genetically engineered mouse models of breast carcinoma: a translational resource for highlighting human breast subtype etiology and developing personalized therapeutic approaches
(Under the direction of Charles M. Perou)

Approximately one in eight women will be diagnosed with breast cancer during their lifetime. While increased public awareness has led to earlier detection of this common disease, a greater understanding of tumor biology has led to the development of many promising therapeutics. A difficult frontier, however, has been identifying the appropriate target population for new drugs as not all breast cancer patients will respond to a particular therapeutic. Currently, approximately five percent of oncology drugs that enter clinical testing are ultimately approved by the US Food and Drug Administration for use. This low success rate reflects not only the difficulty of developing anticancer therapeutics, but also flaws in preclinical testing methodology for selecting the most appropriate cancer patient subset for early clinical testing. With so many patients either not responding or relapsing with the current standard of care, improved personalized therapeutic approaches are greatly needed.

Breast cancer is a heterogeneous disease consisting of multiple intrinsic subtypes. Even though clear clinical and genetic distinctions between the subtypes have been described, the driving mechanisms underlying the initiation, growth and metastasis of breast tumors are under intense investigation to more fully characterize these phenotypes since targeted treatment against specific aberrations promises to be more effective with less systemic side effects. Genetically engineered mouse models are a useful resource for studying mammary

cancers *in vivo* under genetically controlled and immune competent conditions. Identifying murine models with conserved human tumor features will not only facilitate etiology determinations for the intrinsic subtypes, but also serve as a useful preclinical resource for testing the efficacy of new therapeutic approaches. These mice promise to be better predictors of clinical trial success because they resemble tumor biology more closely than other approaches. The work presented here focuses on determining the degree to which current mouse models of breast carcinoma resemble the human disease state and identifies mouse model counterparts for each of the human intrinsic subtypes. If these credentialed mouse models are used for preclinical testing, we anticipate a higher success rate for the development of targeted therapeutics.

I dedicate my research to all those who have been affected by breast cancer; either being diagnosed themselves, knowing a survivor, or celebrating the memory of someone who courageously battled this disease.

ACKNOWLEDGEMENTS

I would like to thank the following individuals who over the course of my lifetime have all contributed to making this dissertation possible. My graduate advisor: **Chuck Perou**, for supporting and encouraging me to tailor my graduate training toward my long-term career goals. My committee members: **Shelley Earp**, **Bill Kaufmann**, **Joel Parker**, **Ned Sharpless**, and **Buddy Weissman**, for your scientific and professional guidance. **Current and past Perou Lab members**, for always being available to lend a helping hand or ear. My undergraduate advisor: **Susan Lord**, for introducing me to the joy of academic research. **Greg Wray**, for helping shape me as a scientist and a person. **My family**, for everything; I would not be here without each and every one of you. My wife: **Lisa Pfefferle**, for supporting and loving me each and every day. My son: **Franklin Pfefferle**, for keeping my life in perspective.

TABLE OF CONTENTS

LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvi
CHAPTER 1: INTRODUCTION	1
A greater understanding of breast cancer biology will identify novel therapeutic targets	1
Personalized drug regimens are the future of cancer treatment	4
Murine models are excellent for translating biological discovery into clinical care	5
References	7
CHAPTER 2: TRANSCRIPTOMIC CLASSIFICATION OF GENETICALLY ENGINEERED MOUSE MODELS OF BREAST CANCER IDENTIFIES HUMAN SUBTYPE COUNTERPARTS	12
Overview	12
Background	12
Results	12
Conclusion	13
Background	13
Results	14
Expression classes of genetically engineered mouse models	14

Comparison of murine class defining gene sets versus human tumor subtypes	23
Conserved tumorigenic pathway signatures identified between human-mouse counterparts	29
Discussion	33
Conclusion	39
Materials and Methods.....	39
Gene expression microarrays.....	39
Murine intrinsic genes and subtypes.....	41
Human and mouse intrinsic gene cocluster	42
Comparison of murine and human expression subtypes.....	42
Conserved pathway gene signatures	43
References	44
CHAPTER 3: LUMINAL PROGENITOR AND FETAL MAMMARY STEM CELL EXPRESSION FEATURES PREDICT BREAST TUMOR RESPONSE TO NEOADJUVANT CHEMOTHERAPY	50
Overview	50
Background	50
Methods.....	50
Results.....	51
Conclusions.....	51
Background	51
Results.....	53
Comparison of human mammary subpopulation transcriptomic datasets	53
Human mammary cell subpopulation enriched gene signatures.....	57

Murine mammary cell subpopulation enriched gene signatures.....	61
LumProg and fMaSC features predict neoadjuvant chemotherapy response.....	65
Discussion	70
Materials and Methods.....	75
Mammary cell subpopulation gene signatures.....	75
Comparison of human and murine normal mammary populations	76
Mammary cell subpopulation centroids.....	76
Chemotherapy response	77
References	79
CHAPTER 4: SECONDARY GENETIC ABERRATION PROFILING OF P53NULL MAMMARY TUMORS HIGHLIGHTS <i>MET</i> DNA AMPLIFICATION AS A GENETIC DRIVER OF MURINE BASAL-LIKE TUMORIGENESIS	84
Overview.....	84
Background	84
Methods.....	84
Results.....	85
Conclusions.....	85
Background	85
Results.....	87
p53null transplant tumors are counterparts for human basal-like and claudin-low subtypes.....	87
Secondary genetic aberration profiling highlights DNA copy number changes as drivers of tumorigenesis	91
2250L (p53null-Luminal ^{Ex}) tumors completely regress with Crizotinib treatment	101

Discussion	101
Materials and Methods.....	104
Gene expression	104
Flow cytometry	105
DNA single nucleotide polymorphisms.....	105
DNA structural variants	106
DNA copy number	106
Crizotinib treatment	107
References	108
CHAPTER 5: THE MMTV-WNT1 MURINE MODEL PRODUCES TWO PHENOTYPICALLY DISTINCT SUBTYPES OF MAMMARY TUMORS WITH UNIQUE CLINICAL OUTCOMES TO EGFR INHIBITORS	112
Overview	112
Background	112
Methods.....	112
Results.....	113
Conclusions.....	113
Background	113
Results.....	116
Wnt1-Early ^{Ex} and Wnt1-Late ^{Ex} tumors have distinct gross pathology and histology traits	116
Wnt1-Early ^{Ex} tumors are enriched for canonical and non-canonical Wnt pathway signatures	120
Wnt1-Early ^{Ex} tumors respond to epidermal growth factor receptor inhibitors	122

Wnt1-Early ^{Ex} and Wnt1-Late ^{Ex} tumors have distinct mammary subpopulation FACS profiles.....	124
Both Wnt1-Late ^{Ex} tumor FACS subpopulations have tumor initiating potential	127
Both Wnt1-Late ^{Ex} tumor FACS subpopulations have activating <i>Hras1</i> mutations.....	129
Discussion	131
Materials and Methods.....	134
Mouse husbandry	134
Gene expression	134
DNA copy number	135
Immunofluorescence	136
Drug treatment	136
Flow cytometry	137
Sanger Sequencing.....	138
References	139
CHAPTER 6: DISCUSSION.....	143
References	151
APPENDIX A: RELATED COAUTHORSHIPS	153
Oncogenic PI3K mutations lead to NF-kB-dependent cytokine expression following growth factor deprivation	153
LKB1/STK11 inactivation leads to expansion of a pro-metastatic tumor sub-population in melanoma.....	154
Comparative oncogenomics implicates the Neurofibromin 1 gene (<i>NF1</i>) as a breast cancer driver	155
Combined PI3K/mTOR and MEK inhibition provides broad anti-tumor activity in faithful murine cancer models	156

Conditional loss of ErbB3 delays mammary gland hyperplasia induced by mutant PIK3CA without affecting mammary tumor latency, gene expression or signaling.....	157
Mutant PIK3CA accelerates HER2-driven transgenic mammary tumors and induces resistance to combinations of anti-HER2 therapies	158
Endothelial-like properties of claudin-low breast cancer cells promote tumor vascular permeability and metastasis	159
c-Myc and Her2 cooperate to drive stem-like phenotype with poor prognosis in breast cancer	160
Expression of miR-200c in claudin-low breast cancer alters stem cell functionality, enhances chemosensitivity and reduces metastatic potential	161
JNK2 prevents luminal cell commitment in normal mammary glands and tumors by inhibiting p53/NOTCH1 and BRCA1 expression	162
APPENDIX B: CURRICULUM VITAE	163
Education	163
Research and Work Experience	163
Publications.....	164
Presentations	166
Teaching and mentoring	167
Certifications.....	168
Outreach.....	168

LIST OF TABLES

Table 1 – Summary of murine models studied	15
Table 2 – Gene set analysis of murine classes	24
Table 3 – Gene set analysis of murine classes and human subtypes	27
Table 4 – Human FACS enriched normal mammary cell subpopulation studies	54
Table 5 – Murine FACS enriched normal mammary cell subpopulation studies	60
Table 6 – Gene set analysis of human and murine cell subpopulations	62
Table 7 – Clinical characteristics of the neoadjuvant chemotherapy treated dataset	66
Table 8 – Univariate logistic regression analysis predicting pathological complete response	67
Table 9 – Multivariate logistic regression analysis predicting pathological complete response	68
Table 10 – DNA amplification status of <i>CUL4A</i> and <i>MET</i> across the TCGA breast cancer dataset	100

LIST OF FIGURES

Figure 1 – Flowchart of murine expression data analysis.....	16
Figure 2 – Murine intrinsic class analysis.....	18
Figure 3 – Murine intrinsic tumor dendrogram by sample	19
Figure 4 – Murine intrinsic tumor dendrogram by mouse model	20
Figure 5 – Murine intrinsic cluster signature according to tumor subtype	22
Figure 6 – Human and murine intrinsic co-cluster	26
Figure 7 – Conserved signaling pathways between human-mouse counterparts	31
Figure 8 – Flowchart of analysis.....	55
Figure 9 – Comparison of mammary subpopulations across studies.....	56
Figure 10 – <i>Homo sapiens</i> enriched gene signatures.....	58
Figure 11 – <i>Mus musculus</i> enriched gene signatures.....	63
Figure 12 – fMaSC enriched gene signatures	69
Figure 13 – Human counterparts of p53null transplant tumors	88
Figure 14 – Molecular pathway signatures and differentiation score.....	89
Figure 15 – Murine p53null tumor datasets.....	93
Figure 16 – Somatic non-silent mutation analysis.....	94
Figure 17 – Chromosome structural variation analysis	95
Figure 18 – DNA copy number analysis.....	97
Figure 19 – <i>MET</i> DNA amplification is a driver of 2250L (p53null-Luminal ^{Ex}) tumors	99
Figure 20 – The MMTV-Wnt1 model produce two classes of mammary tumors with distinct latencies, gross pathology and histology features.....	117

Figure 21 – Wnt1-Early ^{Ex} and Wnt1-Late ^{Ex} have similar DNA copy number landscapes	118
Figure 22 – Wnt-Early ^{Ex} tumors have expression of Wnt associated pathway signatures	121
Figure 23 – Wnt-Early ^{Ex} tumors respond to EGFR inhibitors.....	123
Figure 24 – Wnt-Early ^{Ex} and Wnt1-Late ^{Ex} tumors share features with different normal mammary cell types.....	125
Figure 25 – Both Wnt1-Late ^{Ex} tumor subpopulations have tumor initiating potential.....	128
Figure 26 – Wnt-Late ^{Ex} tumors have activating Hras mutations	130

LIST OF ABBREVIATIONS

aCGH	array comparative genomic hybridization
Cy3	cyanine-3
Cy5	cyanine-5
DNA	deoxyribonucleic acid
D-Score	differentiation score
EGFR	epidermal growth factor receptor
EMT	epithelial-to-mesenchymal transition
ER	estrogen receptor
FACS	fluorescence-activated cell sorting
FDR	false discovery rate
FET	fisher's exact test
GEMM	genetically engineered mouse model
GSA	gene set analysis
H&E	hematoxylin and eosin
HER2	human epidermal growth factor receptor 2
HsEnriched	<i>Homo sapiens</i> enriched
IACUC	institutional animal care and use committee
LumProg	luminal progenitor
MaSC	mammary stem cell
MatureLum	mature luminal
MmEnriched	<i>Mus musculus</i> enriched

MVA	multivariate analysis
PCA	principle component analysis
pCR	pathologic complete response
PR	progesterone receptor
RNA	ribonucleic acid
SAM	significance analysis of microarrays
SNP	single nucleotide polymorphism
SSP	single sample predictor
Str	stroma
SV	structural variant
TCGA	the cancer genome atlas
TNBC	triple negative breast cancer
UVA	univariate analysis

CHAPTER 1: INTRODUCTION

About three quarters of a million American women were diagnosed with new cases of cancer last year [1]. With breast cancer accounting for about thirty percent of these incidences, one in eight women will be affected by this common disease in their lifetime [1]. Increased public awareness and a greater understanding of tumor biology has led to better patient survival rates since the early 1990s, yet this disease is still the second leading cause of cancer related deaths in American women [1]. With so many patients either not responding or relapsing with the current standard of care, the molecular mechanisms underlying breast cancer are under intense investigation to identify new, personalized drug targets that should improve patient outcomes [2, 3].

A greater understanding of breast cancer biology will identify novel therapeutic targets

Breast cancer is a heterogeneous disease consisting of multiple disease subtypes that are referred to as the intrinsic subtypes: luminal A, luminal B, HER2-enriched, basal-like, claudin-low [4-8]. Tumor subtyping can be performed using two complementary gene expression based tests: the PAM50 intrinsic subtyping assay [9] and the claudin-low predictor [6]. Clinically, these subtypes are prognostic [5, 7, 9] and predict sensitivity to specific therapeutics [5, 10]. Luminal A tumors have the best overall survival, while the other subtypes have similar, poor outcomes if untreated [7]. Generally, the luminal A/B subtypes are estrogen receptor positive (ER⁺) and/or progesterone receptor positive (PR⁺) tumors [7]. These subtypes also tend to lack the human

epidermal growth factor receptor 2 (HER2⁻). Luminal B tumors are distinguishable from the luminal A subtype by their faster, more aggressive proliferation rates. HER2-enriched tumors, on the other hand, are typically ER⁻, PR⁻, and HER2⁺, with similar high proliferation and low survival rates as luminal B tumors when untreated [7]. The remaining two subtypes, basal-like and claudin-low, are broadly considered triple negative breast cancer (TNBC) (ER⁻, PR⁻ and HER2⁻) [7]. Basal-like breast cancers have the highest proliferation rates and are the most genetically unstable of all the subtypes [11, 12]. Claudin-low tumors are characterized by low levels of cell adhesion molecules and high levels of inflammatory cells [6]. While targeted therapeutics exist for ER⁺ [13] (luminal A/B [14]) and HER2⁺ [15] (HER2-enriched [14]) breast cancer, targeted treatments for TNBC (basal-like and claudin-low [14]) remain an important unmet clinical need [16]. To address this need, a research emphasis has been placed on determining the molecular drivers of basal-like and claudin-low tumors to identify novel drug targets for these subtypes, which was an important aspect of my thesis work.

Even though clear distinctions between the intrinsic subtypes have been defined, the molecular mechanisms that give rise to breast tumors in general, and the individual intrinsic subtypes, are not fully known. For instance, it is unknown why some tumors of the same subtype respond differently to the same therapeutic treatment [3, 17]. A greater knowledge of the molecular aberrations underlying breast tumors will identify these differences in genetic drivers, which we believe to be the key first step towards personalized drug therapies [3]. Segregating genetic drivers from passenger mutations is difficult due to the inherent heterogeneity of breast tumors and the large number of genetic aberrations seen within a given tumor. Several hypotheses have been described to explain breast cancer heterogeneity. 1) Breast tumors can arise from a number of different mature and/or stem cell types, which may form the basis for the

intrinsic subtypes [18-21]. 2) Similarly, tumors develop through clonal expansion of evolving clones, resulting in intra-tumor pockets of cell colonies containing their own set of aberrations [22]. 3) The tumor microenvironment can interact with progressing cancer cells to affect the overall tumor phenotype [23, 24]. 4) External factors (e.g. environmental, lifestyle, and comorbidity) influence tumor development [25]. All of these factors probably have some role in determining the ultimate phenotype of a progressing breast tumor, but more research is needed to validate the relative contributions of each.

Over the last fifteen years, genomics [12, 26] and transcriptomics [4] have fueled a greater understanding of breast tumor biology [8]. The cancer genome is broadly characterized as unstable. The inability to properly respond to and fix DNA damage leads to the accumulation of small scale mutations [26] (e.g. insertions, deletions) and large scale chromosomal rearrangements [12] (e.g. translocations, aneuploidy). Microarray and sequencing techniques have been developed to identify genomic aberrations that can lead to decreased tumor suppressor function, increased oncogene signaling, or both. Transcriptomics (gene expression analysis) is a popular approach for characterizing tumors because it is easy to measure, provides a rough estimate of corresponding protein levels, and identifies overarching tumor phenotypes. Given the large number of transcriptomic studies, the Broad Institute has created the molecular signatures database (MSigDB), which compiles gene sets/modules from the literature into one place [27]. This database improves on the gene-gene comparison by allowing for the comparison of gene groups. While each of these approaches has improved our knowledge of cancer biology by themselves, studies integrating multiple 'omic approaches have an even greater power of identifying important tumor phenotypes and novel drug targets, with examples being the cancer genome atlas's (TCGA) analysis of glioblastoma and ovarian cancers [28, 29].

Personalized drug regimens are the future of cancer treatment

Developing clinical tests that predict drug response (i.e. companion diagnostics) is an important focus of cancer research. Two clinical trials have brought the use of molecular testing to the forefront of breast cancer research. In these trials, the Oncotype DXTM (TAILORx trial) [30] and MammaPrintTM (MINDACT trial) [31] assays are being used to determine which patients should receive chemotherapy. A recent comparison of Oncotype DXTM with the PAM50 intrinsic subtyping assay identified that the PAM50 risk of recurrence (ROR) as superior to the Oncotype DXTM recurrence score (RS) in endocrine-treated patients with ER-positive, node-negative disease [32]. While the use of molecular testing in the clinic has been groundbreaking, these assays do not identify which drug regimens to prescribe. Clinical assays that determine personalized drug regimens are greatly needed to improve patient survival.

Before personalized drug testing can begin in the clinic, the target population for molecularly targeted drugs needs to be identified. This is a difficult task, but is critical for developing the best predictors of response. Targeted cancer drugs are designed to inhibit specific genetic aberrations, typically kinases or hormone receptors, but not all tumors with the aberration respond to treatment [17]. Additional factors (e.g. coexisting aberrations, cell of origin, microenvironment) are hypothesized to confer sensitivity or resistance to drug treatment [33-36]. Even though studies have defined predictive signatures of response [37-41], most are not reproducible [42, 43]. Many of these predictive signatures were derived under very specific conditions (e.g. using one cell line), so while they might work in a specific limited setting, they typically fail when applied to a broad population of patients [44]. Clinical tumors are highly heterogeneous; therefore, predictive signatures should be tested under heterogeneous conditions and across multiple models. This thesis was in part focused on characterizing the wealth of

mouse models of human breast cancer such that this heterogeneous set of models could be better linked to human breast tumors, which we predict would then help drug testing and development.

Murine models are excellent for translating biological discovery into clinical care

Currently, only ~5% of oncology drugs that enter clinical testing are approved by the FDA for use [45]. This dismal success rate not only reflects the difficulty of developing anticancer therapeutics, but also flaws in preclinical methodology for selecting the most promising drugs to use in clinical trials [43, 46]. Historically, preclinical drug testing has primarily involved a mix of *in vitro* cell line and *in vivo* cell line xenograft experiments. While these studies are a good first step, they do not represent true tumor biology [47]. For instance, the few cell lines available for use have limited biological diversity when compared to primary tumors, and they represent a subset of the original clones with the best growth advantages. These cell line based approaches are unable to accurately represent clinical heterogeneity or an intact microenvironment, and as a result, many of the drugs that pass cell line based preclinical testing fail during later stages.

Even though they have existed for decades, genetically engineered mouse models (GEMMs) have only recently become a more popular system for preclinical drug testing [47]. While cross-species complications do exist, these mice promise to be better predictors of clinical trial success because they resemble tumor biology more closely than xenografts. This is most evident from treatment experiments showing that tumors within a murine model vary in their drug response [48]. This response spectrum resembles the results from human trials more than the typical uni-modal response of cell line-based xenograft studies. These early observations indicate a strong need for murine models in preclinical testing.

Comparative studies between human and murine tumors provide an attractive approach for narrowing the genetic driver candidate list by highlighting conserved features between species [49] and thus focusing attention on the most likely driver genes. While dozens of murine models have been created to study the molecular mechanisms of breast cancer [50-56], the degree to which many of these models recapitulate the human subtypes is largely unknown. Before proper therapeutic comparative studies can be performed, it is essential that human-to-murine tumor counterparts are identified to ensure that the chosen model accurately replicates the genetic alterations and overall phenotypes observed in human tumors [49]. This is especially important for heterogeneous human diseases, such as breast cancer.

Given the advantages of murine models for studying tumorigenesis [47], the following four chapters utilize genetically engineered mouse model of breast carcinoma to simultaneously investigate human tumor etiology and as a tool for preclinical drug testing. Chapter 1, which was published in *Genome Biology* in 2013 [49], analyzes the transcriptomic profiles of 27 murine models to highlight the subset of GEMMs that mimic the human disease state. Chapter 2, which was published in *Breast Cancer Research and Treatment* in 2015 [57], analyzes the transcriptomic profiles of normal mammary cell types to highlight conserved cell features between human-murine subtype counterparts. In addition, this study identifies several gene signatures that predict tumor pathologic complete response sensitivity to neoadjuvant chemotherapy even after controlling for intrinsic subtype, proliferation, and clinical variables. Chapter 3 analyzes the secondary genetic aberrations of p53null mammary transplant murine tumors, highlighting MET as a genetic driver of murine basal-like tumors. Chapter 4 investigates the phenotypic and clinical differences between the MMTV-Wnt1 tumor subtypes, highlighting EGFR as a potential drug target in breast tumors with aberrant Wnt signaling.

REFERENCES

1. AmericanCancerSociety: **Cancer Facts and Figures**. 2015.
2. Curigliano G, Goldhirsch A: **The triple-negative subtype: new ideas for the poorest prognosis breast cancer**. *J Natl Cancer Inst Monogr* 2011, **2011**:108-110.
3. Curigliano G: **New drugs for breast cancer subtypes: Targeting driver pathways to overcome resistance**. *Cancer Treat Rev* 2011, **38**:303-310.
4. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**:747-752.
5. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications**. *Proc Natl Acad Sci U S A* 2001, **98**:10869-10874.
6. 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.
7. Prat A, Perou CM: **Deconstructing the molecular portraits of breast cancer**. *Mol Oncol* 2010.
8. CancerGenomeAtlasNetwork: **Comprehensive molecular portraits of human breast tumours**. *Nature* 2012, **490**:61-70.
9. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, et al: **Supervised risk predictor of breast cancer based on intrinsic subtypes**. *J Clin Oncol* 2009, **27**:1160-1167.
10. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes**. *Clin Cancer Res* 2007, **13**:2329-2334.
11. Rakha EA, Reis-Filho JS, Ellis IO: **Basal-like breast cancer: a critical review**. *J Clin Oncol* 2008, **26**:2568-2581.
12. Weigman VJ, Chao HH, Shabalin AA, He X, Parker JS, Nordgard SH, Grushko T, Huo D, Nwachukwu C, Nobel A, et al: **Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival**. *Breast Cancer Res Treat* 2011.

13. Jordan VC: **Tamoxifen: a most unlikely pioneering medicine.** *Nat Rev Drug Discov* 2003, **2**:205-213.
14. Prat A, Perou CM: **Deconstructing the molecular portraits of breast cancer.** *Mol Oncol* 2011, **5**:5-23.
15. Hynes NE, Lane HA: **ERBB receptors and cancer: the complexity of targeted inhibitors.** *Nat Rev Cancer* 2005, **5**:341-354.
16. Carey L, Winer E, Viale G, Cameron D, Gianni L: **Triple-negative breast cancer: disease entity or title of convenience?** *Nat Rev Clin Oncol* 2010, **7**:683-692.
17. Bates M, Sperinde J, Kostler WJ, Ali SM, Leitzel K, Fuchs EM, Paquet A, Lie Y, Sherwood T, Horvat R, et al: **Identification of a subpopulation of metastatic breast cancer patients with very high HER2 expression levels and possible resistance to trastuzumab.** *Ann Oncol* 2010, **22**:2014-2020.
18. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, et al: **Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.** *Nat Med* 2009, **15**:907-913.
19. Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, Iscove N, Jones S, McKinney S, Emmerman J, et al: **Transcriptome analysis of the normal human mammary cell commitment and differentiation process.** *Cell Stem Cell* 2008, **3**:109-118.
20. Spike BT, Engle DD, Lin JC, Cheung SK, La J, Wahl GM: **A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer.** *Cell Stem Cell* 2012, **10**:183-197.
21. Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, Eaves CJ: **Purification and unique properties of mammary epithelial stem cells.** *Nature* 2006, **439**:993-997.
22. Siegmund KD, Marjoram P, Woo YJ, Tavaré S, Shibata D: **Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers.** *Proc Natl Acad Sci U S A* 2009, **106**:4828-4833.
23. Hatiboglu MA, Kong LY, Wei J, Wang Y, McEnery KA, Fuller GN, Qiao W, Davies MA, Priebe W, Heimberger AB: **The tumor microenvironment expression of p-STAT3 influences the efficacy of cyclophosphamide with WP1066 in murine melanoma models.** *Int J Cancer* 2012, **131**:8-17.
24. Bissell MJ, Radisky DC, Rizki A, Weaver VM, Petersen OW: **The organizing principle: microenvironmental influences in the normal and malignant breast.** *Differentiation* 2002, **70**:537-546.

25. Hurria A: **Embracing the complexity of comorbidity.** *J Clin Oncol* 2011, **29**:4217-4218.
26. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, et al: **The consensus coding sequences of human breast and colorectal cancers.** *Science* 2006, **314**:268-274.
27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**:15545-15550.
28. Consortium TCGA: **Integrated genomic analyses of ovarian carcinoma.** *Nature* 2011, **474**:609-615.
29. Consortium TCGA: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061-1068.
30. Zujewski JA, Kamin L: **Trial assessing individualized options for treatment for breast cancer: the TAILORx trial.** *Future Oncol* 2008, **4**:603-610.
31. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ: **Clinical application of the 70-gene profile: the MINDACT trial.** *J Clin Oncol* 2008, **26**:729-735.
32. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, Ferree S, Storhoff J, Schaper C, Cuzick J: **Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy.** *J Clin Oncol* 2013, **31**:2783-2790.
33. Chaft JE, Arcila ME, Paik PK, Lau C, Riely GJ, Pietanza MC, Zakowski MF, Rusch V, Sima CS, Ladanyi M, Kris MG: **Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma-rationale for comprehensive mutation profiling.** *Mol Cancer Ther* 2011, **11**:485-491.
34. Yuan TL, Cantley LC: **PI3K pathway alterations in cancer: variations on a theme.** *Oncogene* 2008, **27**:5497-5510.
35. Kuo YW, Wu SG, Ho CC, Shih JY: **Good response to gefitinib in lung adenocarcinoma harboring coexisting EML4-ALK fusion gene and EGFR mutation.** *J Thorac Oncol* 2010, **5**:2039-2040.
36. Britton KM, Eyre R, Harvey IJ, Stemke-Hale K, Browell D, Lennard TW, Meeson AP: **Breast Cancer, Side Population cells and ABCG2 expression.** *Cancer Lett* 2012.

37. Augustine CK, Jung SH, Sohn I, Yoo JS, Yoshimoto Y, Olson JA, Jr., Friedman HS, Ali-Osman F, Tyler DS: **Gene expression signatures as a guide to treatment strategies for in-transit metastatic melanoma.** *Mol Cancer Ther* 2010, **9**:779-790.
38. Bild AH, Parker JS, Gustafson AM, Acharya CR, Hoadley KA, Anders C, Marcom PK, Carey LA, Potti A, Nevins JR, Perou CM: **An integration of complementary strategies for gene-expression analysis to reveal novel therapeutic opportunities for breast cancer.** *Breast Cancer Res* 2009, **11**:R55.
39. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, et al: **Oncogenic pathway signatures in human cancers as a guide to targeted therapies.** *Nature* 2006, **439**:353-357.
40. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, Ward L, Koo JH, Gopalakrishnan V, Zhu Y, et al: **Oncogenic pathway combinations predict clinical prognosis in gastric cancer.** *PLoS Genet* 2009, **5**:e1000676.
41. Wu CJ, Cai T, Rikova K, Merberg D, Kasif S, Steffen M: **A predictive phosphorylation signature of lung cancer.** *PLoS One* 2009, **4**:e7994.
42. Borst P, Wessels L: **Do predictive signatures really predict response to cancer chemotherapy?** *Cell Cycle* 2010, **9**.
43. Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature* 2012, **483**:531-533.
44. Weigelt B, Pusztai L, Ashworth A, Reis-Filho JS: **Challenges translating breast cancer gene signatures into the clinic.** *Nat Rev Clin Oncol* 2011, **9**:58-64.
45. Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nat Rev Drug Discov* 2004, **3**:711-715.
46. Hutchinson L, Kirk R: **High drug attrition rates--where are we going wrong?** *Nat Rev Clin Oncol* 2011, **8**:189-190.
47. Sharpless NE, Depinho RA: **The mighty mouse: genetically engineered mouse models in cancer drug development.** *Nat Rev Drug Discov* 2006, **5**:741-754.
48. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L, Karginova O, Jordan J, Combest A, Bridges A, et al: **Predicting drug responsiveness in human cancers using genetically engineered mice.** *Clin Cancer Res* 2013, **19**:4889-4899.
49. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzuayus MI, Brown M, Egan SE, Wahl GM, et al: **Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts.** *Genome Biol* 2013, **14**:R125.

50. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, et al: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors.** *Genome Biol* 2007, **8**:R76.
51. Chan SR, Vermi W, Luo J, Lucini L, Rickert C, Fowler AM, Lonardi S, Arthur C, Young LJ, Levy DE, et al: **STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas.** *Breast Cancer Res* 2012, **14**:R16.
52. Guy CT, Cardiff RD, Muller WJ: **Activated neu induces rapid tumor progression.** *J Biol Chem* 1996, **271**:7673-7678.
53. Husler MR, Kotopoulos KA, Sundberg JP, Tennent BJ, Kunig SV, Knowles BB: **Lactation-induced WAP-SV40 Tag transgene expression in C57BL/6J mice leads to mammary carcinoma.** *Transgenic Res* 1998, **7**:253-263.
54. Pond AC, Herschkowitz JI, Schwertfeger KL, Welm B, Zhang Y, York B, Cardiff RD, Hilsenbeck S, Perou CM, Creighton CJ, et al: **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.
55. Zhang X, Podsypanina K, Huang S, Mohsin SK, Chamness GC, Hatsell S, Cowin P, Schiff R, Li Y: **Estrogen receptor positivity in mammary tumors of Wnt-1 transgenic mice is influenced by collaborating oncogenic mutations.** *Oncogene* 2005, **24**:4220-4231.
56. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, Knezevic J, Greene SB, Darr D, Troester MA, et al: **Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells.** *Proc Natl Acad Sci U S A* 2011, **109**:2778-2783.
57. Pfefferle AD, Spike BT, Wahl GM, Perou CM: **Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant chemotherapy.** *Breast Cancer Res Treat* 2015, **149**:425-437.

CHAPTER 2: TRANSCRIPTOMIC CLASSIFICATION OF GENETICALLY ENGINEERED MOUSE MODELS OF BREAST CANCER IDENTIFIES HUMAN SUBTYPE COUNTERPARTS¹

OVERVIEW

Background

Human breast cancer is a heterogeneous disease consisting of multiple molecular subtypes. Genetically engineered mouse models are a useful resource for studying mammary cancers *in vivo* under genetically controlled and immune competent conditions. Identifying murine models with conserved human tumor features will facilitate etiology determinations, highlight the effects of mutations on pathway activation, and should improve preclinical drug testing.

Results

Transcriptomic profiles of 27 murine models of mammary carcinoma and normal mammary tissue were determined using gene expression microarrays. Hierarchical clustering analysis identified 17 distinct murine subtypes. Cross-species analyses using three independent human breast cancer datasets identified eight murine classes that resemble specific human breast cancer subtypes. Multiple models were associated with human basal-like tumors including TgC3(1)-*Tag*, TgWAP-*Myc* and *Trp53*^{-/-}. Interestingly, the

¹This chapter previously appeared as an article in *Genome Biology*. The original citation is as follows: Pfefferle AD *et al*, “Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts”, *Genome Biology* 2013

TgWAPCre-*Etv6* model mimicked the HER2-enriched subtype, a group of human tumors without a murine counterpart in previous comparative studies. Gene signature analysis identified hundreds of commonly expressed pathway signatures between linked mouse and human subtypes, highlighting potentially common genetic drivers of tumorigenesis.

Conclusions

This study of murine models of breast carcinoma encompasses the largest comprehensive genomic dataset to date to identify human-to-mouse disease subtype counterparts. Our approach illustrates the value of comparisons between species to identify murine models that faithfully mimic the human condition and indicates that multiple mouse models are needed to represent the diversity of human breast cancers. The reported *trans*-species associations should guide model selection during preclinical study design to ensure appropriate representatives of human disease subtypes are used.

BACKGROUND

Breast cancer is the second leading cause of cancer related deaths in American women [1]. While increased public awareness has led to earlier detection, a greater understanding of tumor biology has led to the development of many promising therapeutics [2, 3]. A difficult frontier, however, has been identifying the appropriate target population for new drug(s) as not all breast cancer patients will respond to a particular therapeutic. Currently, only approximately five percent of oncology drugs that enter clinical testing are ultimately approved by the US Food and Drug Administration for use [4]. This low success rate reflects not only the difficulty of developing anticancer therapeutics, but also identifies

flaws in preclinical testing methodology for selecting the most appropriate cancer patient subset for early clinical testing [5, 6].

Numerous murine models of breast cancer have been created to mimic the genetic aberrations found in human tumors [7-30]. Historically, each model has been analyzed independent of other models, which complicates effective comparisons with human tumors. However, when multiple models are consolidated into a single dataset, there is increased sensitivity to detect features that are conserved with the human disease state [31, 32]. Identifying murine models that faithfully mimic specific human breast cancer subtypes [33-35] is an important need for the proper interpretation of mouse model results, and thus, for translating preclinical findings into effective human clinical trials [36]. To address this need, we used a transcriptomic approach to profile tumors from 27 different genetically engineered mouse models (GEMMs). We define and characterize 17 distinct murine subtypes of mammary carcinoma (referred to as classes herein to distinguish them from the human subtypes), which we compare to three human breast tumor datasets comprising over 1700 patients to determine which GEMM classes resemble specific human breast cancer subtypes.

RESULTS

Expression classes of genetically engineered mouse models

As the genetic aberrations of human breast cancers have been elucidated, murine models have been created to investigate the specific role that these genes/proteins have on tumor phenotype. Since our initial comparative genomics study of 14 mouse models and normal mammary tissue [31], the number of breast cancer GEMMs in our database has roughly doubled to 27 (Table 1). To compare the transcriptomic diversity of these GEMMs,

A.				B.		
Tumor Model	Strain	Promoter	Transgene	Ref	Primarily Found in Murine Class(es):	Intramodel Variation
<i>Brg1</i> ^{+/+}	Mixed		<i>Brg1</i> heterozygous	7	Squamous-like ^{Ex} (4/12); Erbb2-like ^{Ex} (3/12); 3 Others	Heterogeneous
Normal Mammary - Lactating	FVB		Normal lactating mammary tissue		Normal-like ^{Ex} (2/2)	Homogeneous
<i>p18</i> ^{-/-}	BALB/c		<i>p18</i> homozygous null	8	Erbb2-like ^{Ex} (5/9); Normal-like ^{Ex} (2/9); Squamous-like ^{Ex} (1/9)	Heterogeneous
<i>Pik3ca</i> -H1047R	FVB	MMTV	<i>Pik3ca</i> H1047R mutation overexpression	9	Class14 ^{Ex} (5/12); Squamous-like ^{Ex} (5/12); 2 Others	Semi-Homogeneous
<i>Rb</i> ^{-/-}	Mixed		<i>Rb</i> homozygous null	10	Erbb2-like ^{Ex} (4/10); Neu ^{Ex} (1/10); 3 Others	Heterogeneous
<i>Stat1</i> ^{-/-}	C57BL/6J		<i>Stat1</i> homozygous null	11	Stat1 ^{Ex} (7/7)	Homogeneous
TgMMTV-Aib1	FVB	MMTV	<i>Aib1</i> overexpression	12	Erbb2-like ^{Ex} (4/9); Myc ^{Ex} (2/9); 2 Others	Heterogeneous
TgMMTV-Atx	FVB	MMTV	<i>Atx</i> overexpression	13	Class14 ^{Ex} (3/5); Squamous-like ^{Ex} (1/5); 1 Other	Semi-Homogeneous
TgMMTV-Fgf3	FVB	MMTV	<i>Fgf3</i> overexpression	14	Erbb2-like ^{Ex} (2/5); Normal-like ^{Ex} (2/5); Wnt1-Late ^{Ex} (1/5)	Semi-Homogeneous
TgMMTV-Hras	FVB	MMTV	<i>Hras</i> overexpression	15	Neu ^{Ex} (5/8); Class8 ^{Ex} (2/8)	Semi-Homogeneous
TgMMTV-Lpa	FVB	MMTV	<i>Lpa1</i> , <i>Lpa2</i> , or <i>Lpa3</i> overexpression	12	Normal-like ^{Ex} (6/15); Claudin-low ^{Ex} (3/15); 3 Others	Heterogeneous
TgMMTV-Myc	FVB	MMTV	<i>Myc</i> overexpression	15	Myc ^{Ex} (4/5); Class8 ^{Ex} (1/5)	Homogeneous
TgMMTV-Wnt1, iFgfr	FVB	MMTV	<i>Wnt1</i> overexpression, inducible <i>Fgfr1</i> or <i>Fgfr2</i>	16	Wnt1-Early ^{Ex} (7/12)	Homogeneous
TgWapCre-Etv6	Mixed	WAP	<i>Etv6-Ntrk3</i> fusion gene overexpression	17	Erbb2-like ^{Ex} (12/12)	Homogeneous
<i>Brca1</i> ^{+/+} , <i>Trp53</i> ^{+/+} , Irradiated	BALB/c		<i>Brca1</i> and <i>Trp53</i> heterozygous, Irradiated	18	p53null-Basal ^{Ex} (6/7); Wnt1-Early ^{Ex} (1/7)	Homogeneous
DMBA-Induced	FVB		DMBA treated	19	Squamous-like ^{Ex} (4/11); Claudin-low ^{Ex} (3/11); 3 Others	Heterogeneous
Normal Mammary	Mixed		Normal mammary tissue		Normal-like ^{Ex} (16/16)	Homogeneous
TgC3(1)-Tag	FVB	C3	SV40 large T antigen	20	C3Tag ^{Ex} (28/30); Claudin-low ^{Ex} (2/30)	Homogeneous
TgMMTV-Cre <i>Brca1</i> ^{loxP/loxP} , <i>Trp53</i> ^{+/+}	C57BL/6J	MMTV	<i>Brca1</i> flox <i>Trp53</i> heterozygous	21	p53null-Basal ^{Ex} (4/10); Claudin-low ^{Ex} (3/10); 1 Other	Heterogeneous
TgMMTV-Neu	FVB	MMTV	Rat <i>Her2</i> overexpression	22	Neu ^{Ex} (25/28); Normal-like ^{Ex} (2/28); 1 Other	Homogeneous
TgMMTV-PyMT	FVB	MMTV	<i>Py-MT</i> overexpression	23	PyMT ^{Ex} (9/17); Class3 ^{Ex} (1/17)	Homogeneous
TgMMTV-Wnt1	FVB	MMTV	<i>Wnt1</i> overexpression	24	Wnt1-Early ^{Ex} (15/25); Wnt1-Late ^{Ex} (7/25); 3 Others	Semi-Homogeneous
TgWap-Int3	FVB	WAP	<i>Notch4</i> overexpression	25	WapInt3 ^{Ex} (6/7); Class3 ^{Ex} (1/7)	Homogeneous
TgWap-Myc	FVB	WAP	<i>cMyc</i> overexpression	26	Myc ^{Ex} (18/21); Class8 ^{Ex} (3/21)	Homogeneous
TgWap-T121	Mixed	WAP	<i>pRb</i> , <i>p107</i> , <i>p130</i> inactivation	27	Erbb2-like ^{Ex} (3/6); Class3 ^{Ex} (2/6); Claudin-low ^{Ex} (1/6)	Semi-Homogeneous
TgWap-T121, <i>Trp53</i> ^{+/+}	B6D2F1	WAP	<i>pRb</i> , <i>p107</i> , <i>p130</i> inactivation, <i>Trp53</i> het	27	C3Tag ^{Ex} (1/1)	NA
TgWap-Tag	C57BL/6J	WAP	SV40 large T antigen	28	C3Tag ^{Ex} (4/4)	Homogeneous
<i>Trp53</i> ^{-/-}	BALB/c		<i>Trp53</i> homozygous null	29	p53null-Luminal ^{Ex} (27/58); p53null-Basal ^{Ex} (15/58); 5 Others	Heterogeneous
<i>Trp53</i> ^{+/+} , Irradiated	BALB/c		<i>Trp53</i> heterozygous, Irradiated	30	p53null-Basal ^{Ex} (4/8); Claudin-low ^{Ex} (2/8); 2 Others	Heterogeneous

Table 1: Summary of murine models studied

A complete list of all GEMMs used. The bottom 15 models/normal mammary were studied by Herschkowitz *et al.* 2007. C3(1), 5' flanking region of the C3(1) component of the rat prostate steroid binding protein. MMTV, mouse mammary tumor virus. WAP, whey acidic protein.

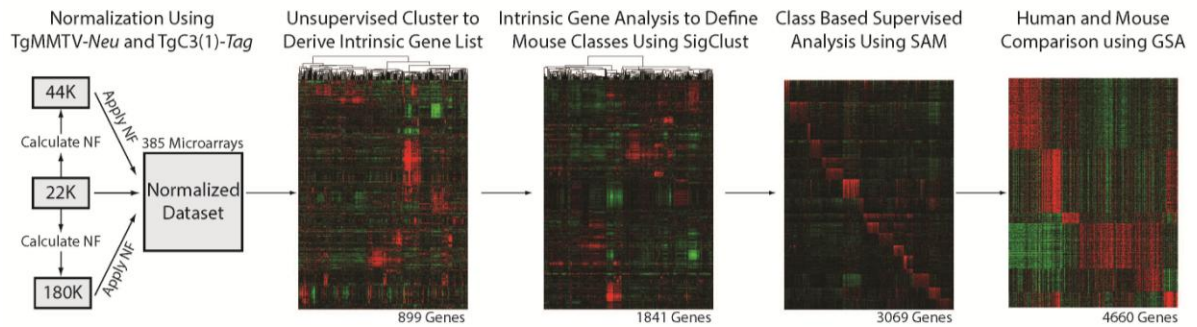


Figure 1: Flowchart of murine expression data analysis

Agilent microarrays from three different platforms were normalized and combined together to create a single murine expression dataset. Next, an unsupervised cluster analysis using variably expressed genes was performed to define a murine ‘intrinsic gene list’. Third, this intrinsic list was used as part of a supervised cluster analysis to objectively define murine subtypes/classes. Fourth, class based supervised analyses were used to define murine class specific lists (genes and pathways). Finally, supervised comparative analysis between human subtypes and mouse classes was used to identify and characterize human-mouse counterparts. Key: NF – normalization factor

global gene expression measurements from 356 unique murine tumors and 16 normal murine mammary samples were analyzed using Agilent microarrays (see Table 1A, Figure 1). Using this larger and more diverse murine dataset, a new mouse ‘intrinsic gene list’ was derived to identify genes associated with all 27 models. As expected, many of the genes from the previous intrinsic gene list were also present in the updated list. After filtering for genes found in both datasets, 76.5% (500/654) of the intrinsic probes from Herschkowitz *et al* 2007 were again included within the new intrinsic list of 1855 probes, which represents 1841 genes.

To determine if new murine subtypes/classes exist in this expanded dataset, SigClust analysis [37] was performed using supervised hierarchical clustering of the 385 murine microarrays and the intrinsic 1855 probe list (Figure 2). Murine ‘classes’ were defined as having at least five tumors with a SigClust p-value ≤ 0.01 . Using these criteria, 17 murine classes were identified with 94% (363/385) of tumors being included within one of these classes (Figure 2B and Figure 3). The name for each class was determined based upon the major model contributor (e.g. Myc^{Ex}), the major biological feature (e.g. Squamous-like^{Ex}), or both (e.g. p53null-Basal^{Ex}), with the ‘^{Ex}’ designation used to denote that this is an expression-based class. As previously observed [31], the *Brca1*^{+/-} *Trp53*^{+/-} irradiated, TgC3(1)-*Tag*, TgMMTV-*Neu*, TgWAP-*Int3*, TgWAP-*Myc*, and TgWAP-*Tag* murine models have ‘homogeneous’ gene expression patterns in this dataset; here, a model was considered ‘homogenous’ if $\geq 80\%$ of tumors from that GEMM were found within a single expression-defined class (Table 1B and Figure 4). Many of the newest models also showed homogeneous gene expression patterns including *Stat1*^{-/-}, TgMMTV-*Myc*, TgMMTV-*Wnt1/iFGFR2*, and TgWAPCre-*Etv6*.

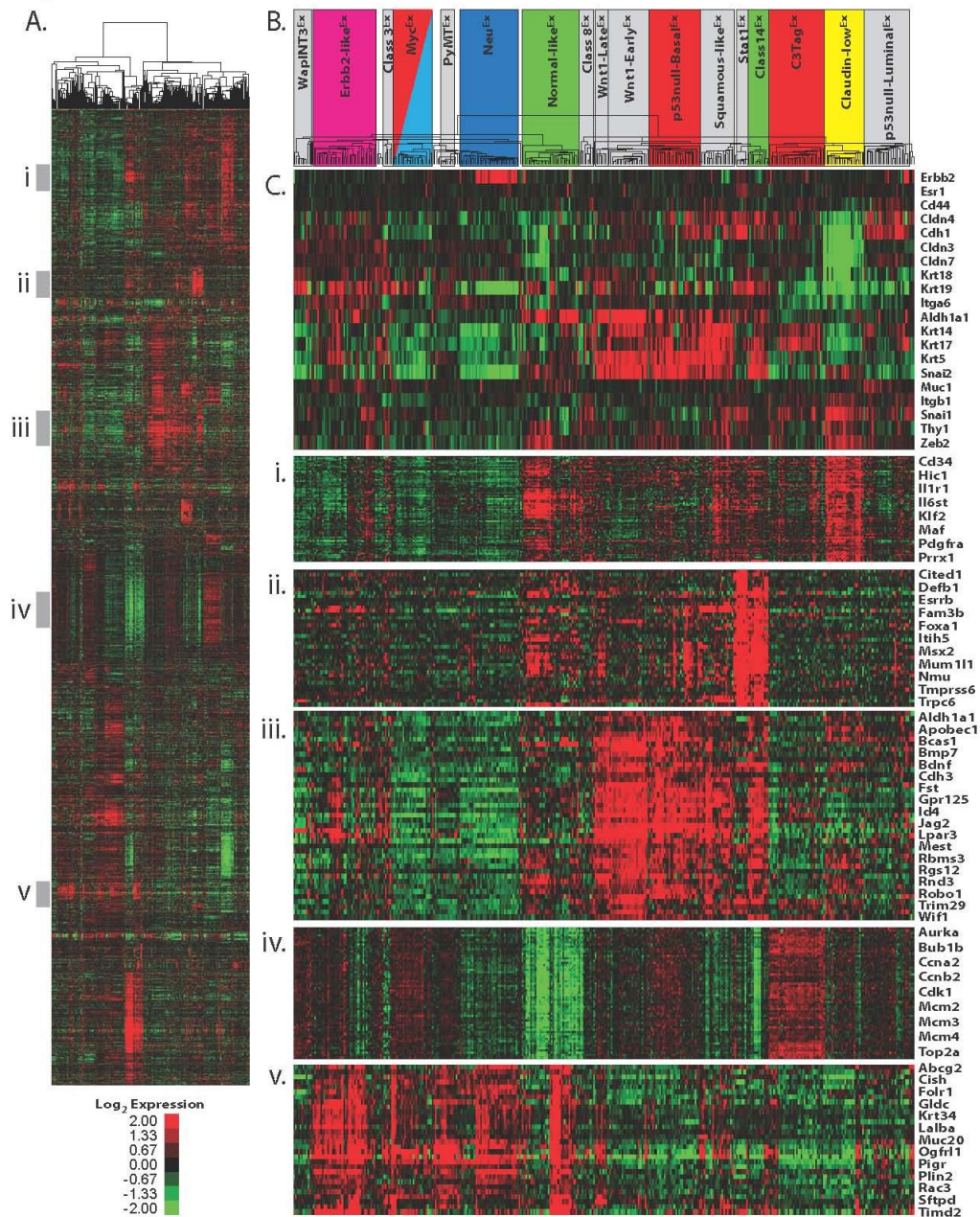


Figure 2: Murine intrinsic class analysis

A. Supervised cluster using the newly derived murine intrinsic gene list and all murine arrays in the dataset. Roman numerals next to the gray bars correspond to the enlarged regions in parts i-v. **B.** Dendrogram of the cluster from part A with the murine classes identified by SigClust highlighted. Classes with colored boxes have been determined to be human expression-based subtype counterparts. **C.** Breast cancer genes and individual cell lineage marker expression profiles. **i.** Claudin-low gene cluster **ii.** Luminal gene cluster **iii.** Basal gene cluster **iv.** Proliferation gene cluster **v.** Lactating gene cluster.

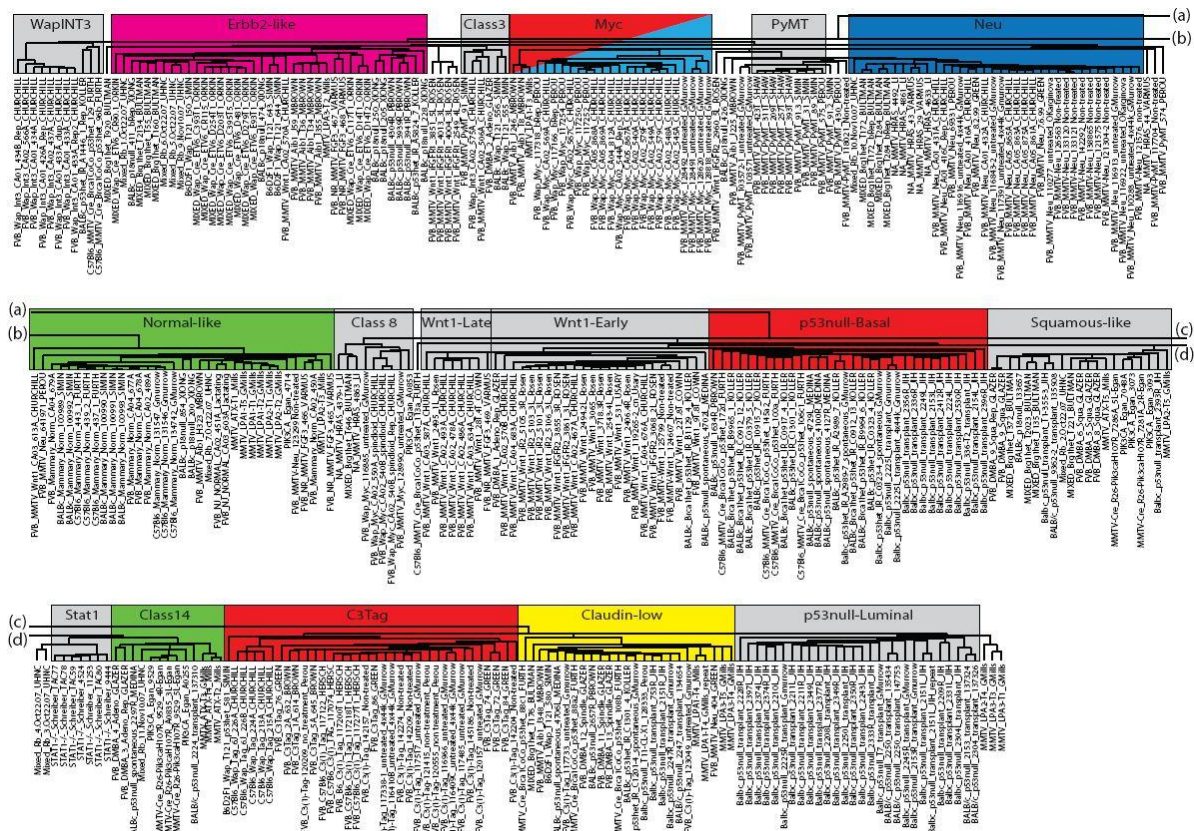


Figure 3: Murine intrinsic tumor dendrogram by sample.
Clustering location of all murine tumors from Figure 2B.

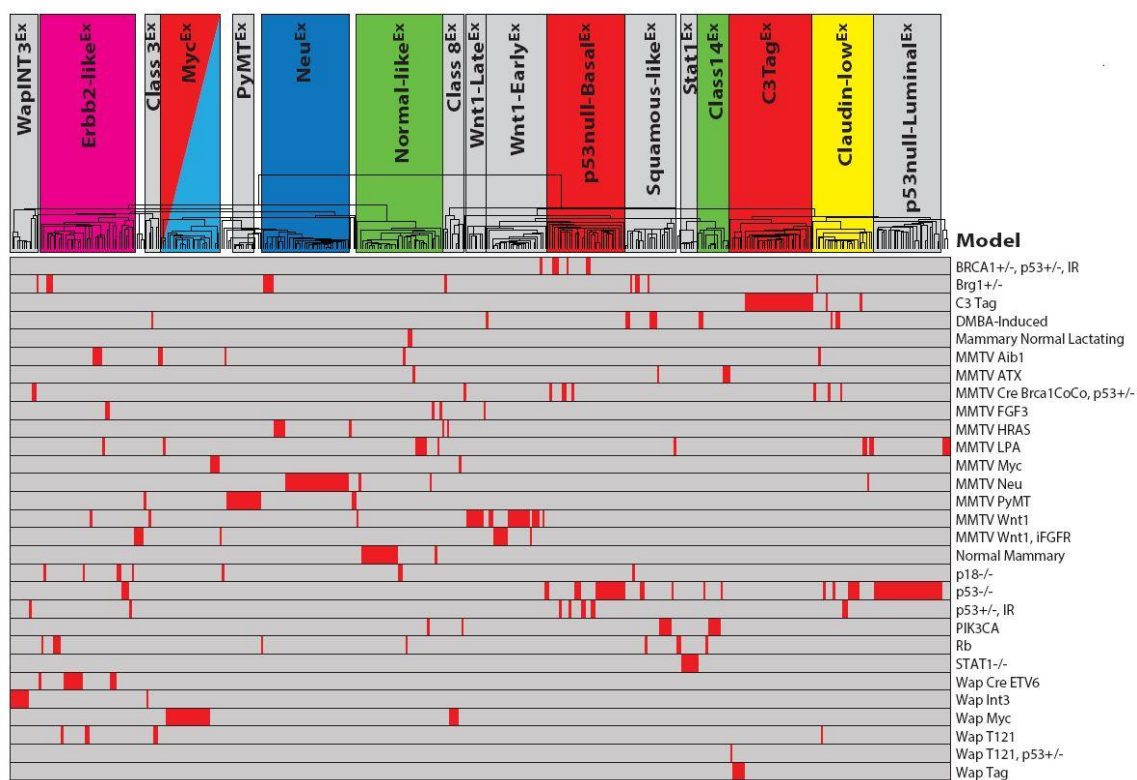


Figure 4: Murine intrinsic tumor dendrogram by mouse model.
Clustering location of all tumors from Figure 2B as displayed by their mouse model.

Other models showed a ‘semi-homogeneous’ gene expression pattern, defined as \geq 80% of tumors from a single GEMM being found within two classes. These included *Pik3ca*-H1047R, TgMMTV-*Atx*, TgMMTV-*Fgf3*, TgMMTV-*Hras*, TgWap-*Tl21*, and TgMMTV-*Wnt1*. Interestingly while maintaining the TgMMTV-*Wnt1* mouse colony, it was observed that there might be two types of tumors based on latency, namely early and late arising tumors. This observation was also reflected in the two TgMMTV-*Wnt1* expression classes that also differed by median tumor latency: Wnt1-Early^{Ex} (8.8 weeks) and Wnt1-Late^{Ex} (22.2 weeks) (Wilcoxon Rank Sum p-value < 0.001). Lastly, about 40% of MMTV provirus driven Wnt1 tumors have cooperative activation of FGF signaling [38], a phenotype that is known to decrease tumor latency [16], and consistent with this, 88% (7/8) of TgMMTV-*Wnt1/iFgfr2* tumors in our dataset were also classified as Wnt1-Early^{Ex}.

The remaining models had ‘heterogeneous’ gene expression patterns, which were defined as no two classes containing at least 80% of the tumors analyzed: *Brg1*^{+/-} (five classes), DMBA-induced (five), *p18*^{-/-} (three), *Rb1*^{-/-} (five), TgMMTV-*Aib1* (four), TgMMTV-Cre *Brca*^{Co/Co} *Trp53*^{+/-} (three), TgMMTV-*Lpa* (four), *Trp53*^{-/-} (seven), and *Trp53*^{+/-} irradiated (four). Similar to recent reports [32], the *Trp53*^{-/-} model (which is distinct from the *Trp53*^{+/-} irradiated model) was primarily defined by three murine classes in this analysis: p53null-luminal^{Ex} (27/58), p53null-basal^{Ex} (15/58), and Claudin-low^{Ex} (7/58).

To begin investigating the defining features of these classes, a comparison of selected cell lineage markers was performed (Figure 2C). Several mouse classes highly expressed luminal cell markers (e.g. *ErbB2*, *Esr1*, *Krt18*, and/or *Krt19*), including ErbB2-like^{Ex}, PyMT^{Ex}, Neu^{Ex}, Myc^{Ex}, and Stat1^{Ex}. Other classes expressed basal cell cytokeratins (e.g. *Krt5*, *Krt14* and/or *Krt17*) including Wnt1-Late^{Ex}, Wnt1-Early^{Ex}, p53null-Basal^{Ex},

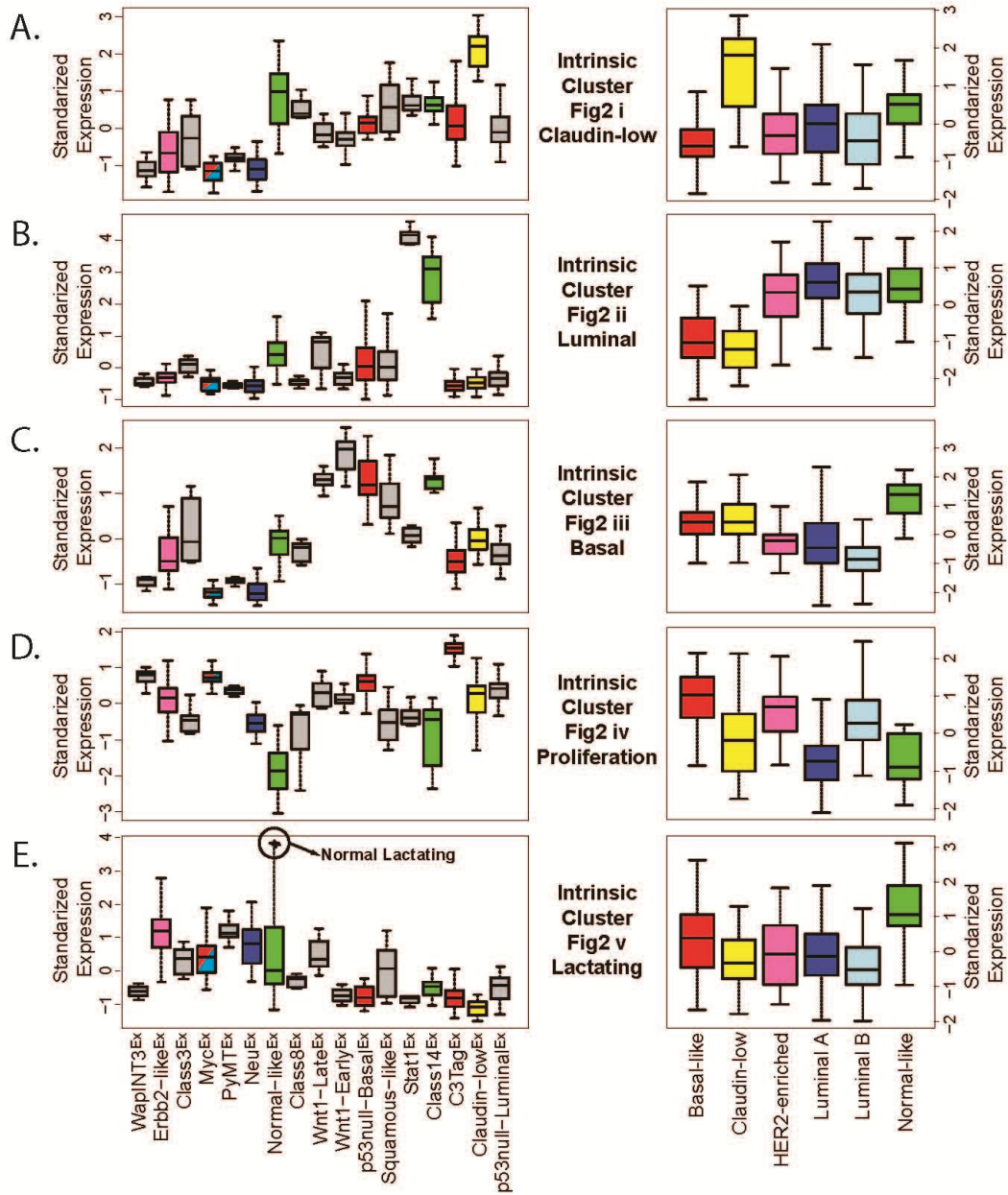


Figure 5: Murine intrinsic cluster signatures according to tumor subtype

Standardized, average expression values for the dominant individual gene clusters from Figure 2i-v are shown according to the murine classes (left panels) and the human subtypes (right panels) using the human UNC308 human breast cancer dataset. **A.** Murine claudin-low subtype defining gene set. **B.** Murine luminal subtype gene set. **C.** Murine basal-like subtype gene set. **D.** Murine proliferation-associated gene set. **E.** Murine lactation associated gene set.

Squamous-like^{Ex}, Class14^{Ex}, and C3Tag^{Ex}. As identified previously [31], a murine Claudin-low^{Ex} class was observed to be characterized by low expression of multiple cell adhesion genes (*Cldn3*, *Cldn4*, and *Cldn7*) and high expression of epithelial-to-mesenchymal transition (EMT) genes (*Snail* and *Zeb2*), similar to the human claudin-low subtype [34].

Comparison of murine class defining gene sets versus human tumor subtypes

To specifically compare murine classes to human breast cancer subtype features, each murine class defining signature (Figure 2i-v) was tested for differential expression across the human subtypes using the UNC308 dataset (Figure 5A-E) [34]. For example, the high expression signature that defines the murine Claudin-low^{Ex} class (Figure 2i, including *Hic1*, *Il6st*, *Klf2*, *Maf*, *Pdgfra*, *Prrx1*, *Snail*) was also the most highly expressed in human claudin-low tumors (Figure 5A).

Figure 2ii shows genes that are highly expressed in the newly identified Stat1^{Ex} and Class14^{Ex} murine classes, which show luminal characteristics (e.g. *Foxa1*, *Esrrb*) and are the most highly expressed in human luminal A tumors (Figure 5B). While most of the GEMMs in this dataset are considered estrogen receptor (ER) negative, murine models comprising these two classes (*Stat1*^{-/-} and *Pik3ca*-H1047R, respectively) were often ER α ⁺ [9, 11], and these data suggest that they overall have a ‘luminal’ expression profile. Interestingly, these classes cluster independent from the previously defined murine luminal models, TgMMTV-*Neu* and TgMMTV-*PyMT*. Consistent with the individual cell lineage marker analysis, the Wnt1-Late^{Ex}, Wnt1-Early^{Ex}, p53null-Basal^{Ex}, Squamous-like^{Ex}, and Class14^{Ex} murine classes express a basal-like gene signature (Figure 2iii). As in human tumors, a proliferation signature (Figure 2iv) further distinguishes these murine classes, with highest expression in

Class	Herschkowitz et al 2007 Class										Predicted Herschkowitz Class
	I-Normal	II-Claudin-low	III-DMBAwnt	IV-BRCAwnt	V-p53null	VI-NeuPyMT	VII-Myc	VIII-WapINT3	IX-WapTag	X-C3Tag	
WapINT3 ^{Ex}	-	-	-	-	-	-	0.324	<1.0e-4	0.352	0.405	WapINT3
ErbB2-like ^{Ex}	-	-	-	0.444	0.059	0.169	0.234	-	0.046	-	
Class3 ^{Ex}	0.275	0.457	0.162	-	-	-	0.288	0.131	-	-	
Myc ^{Ex}	-	-	-	-	-	-	<1.0e-4	-	-	-	Myc
PyMT ^{Ex}	-	-	-	-	-	0.009	0.092	-	0.464	-	NeuPyMT
Neu ^{Ex}	0.322	-	-	-	-	<1.0e-4	-	0.363	-	-	NeuPyMT
Normal-like ^{Ex}	<1.0e-4	0.286	0.289	-	-	0.308	-	0.349	-	-	Normal
Class8 ^{Ex}	0.375	0.104	-	-	-	0.386	0.323	0.337	-	-	
Wnt1-Late ^{Ex}	-	-	<1.0e-4	0.275	-	-	0.357	0.042	-	-	DMBAwnt
Wnt1-Early ^{Ex}	-	-	0.020	0.005	-	0.475	0.344	0.221	-	0.410	BRCA/DMBAwnt
p53null-Basal ^{Ex}	-	0.275	0.089	<1.0e-4	0.068	-	-	-	0.339	0.382	BRCAwnt
Squamous-like ^{Ex}	0.327	0.104	0.003	0.325	-	-	-	0.353	-	-	DMBAwnt
Stat1 ^{Ex}	0.382	0.232	0.164	0.421	0.464	-	-	-	0.431	0.337	
Class14 ^{Ex}	0.366	0.376	0.001	0.278	-	-	-	-	-	-	DMBAwnt
C3Tag ^{Ex}	-	-	-	-	0.454	-	-	-	0.003	<1.0e-4	C3/WapTag
Claudin-low ^{Ex}	-	<1.0e-4	-	0.483	0.330	-	-	-	0.464	-	Claudin-low
p53null-Luminal ^{Ex}	-	0.286	-	0.298	0.001	-	0.239	-	0.125	0.339	p53null

Table 2: Gene set analysis of murine classes

Displayed are the p-values for the gene set analysis comparison of each murine class versus each murine class described in Herschkowitz *et al* 2007. Empty boxes are trending associations, while filled boxes are significant associations ($p < 0.05$, $FDR < 0.1$).

murine C3Tag^{Ex} and human basal-like tumors, and lowest expression in normal tissues from both species. This finding is likely due to the loss of RB1 function in both human basal-like [39, 40] and TgC3(1)-*Tag* murine tumors (due to T-antigen expression). Lastly, Figure 2v highlights a gene cluster that is highly expressed in several murine classes including Erbb2-like^{Ex}, PyMT^{Ex}, and Neu^{Ex}; this signature was lower in normal mammary tissue, but highly expressed in the two lactating mammary samples (Figure 5E). Consistent with this observation, many of the genes in this signature are involved in alveolar function (e.g. *Abcg2*, *Folr1*, and *Lalba*).

For the dual purpose of validating our new classification system and for investigating the degree of diversity in our expanded dataset, the murine classes defined here were compared to those from Herschkowitz *et al* 2007 [31] using gene set analysis (GSA) (Table 2). The majority of the Herschkowitz *et al* 2007 classes had one-to-one matching counterparts to those described here; however, two previous groups (IX-WapTag and X-C3Tag) were combined into a single class in our dataset (C3Tag^{Ex}). Importantly, several of the 17 murine classes defined here were not present within the ten classes of Herschkowitz *et al* 2007 (Erbb2-like^{Ex}, Class3^{Ex}, Class8^{Ex}, and Stat1^{Ex}), almost all of which were populated by GEMMs that were new to this study.

Given the discovery of novel murine classes, it was of great interest to determine the degree to which this expanded murine dataset might better encompass the molecular diversity of the human subtypes. To directly compare tumors across species, this mouse and the previously published UNC308 human datasets were normalized into a single expression dataset and hierarchical clustered using a combined mouse and human [41] intrinsic gene list (Figure 6). While technical differences between the two datasets (e.g. different microarray

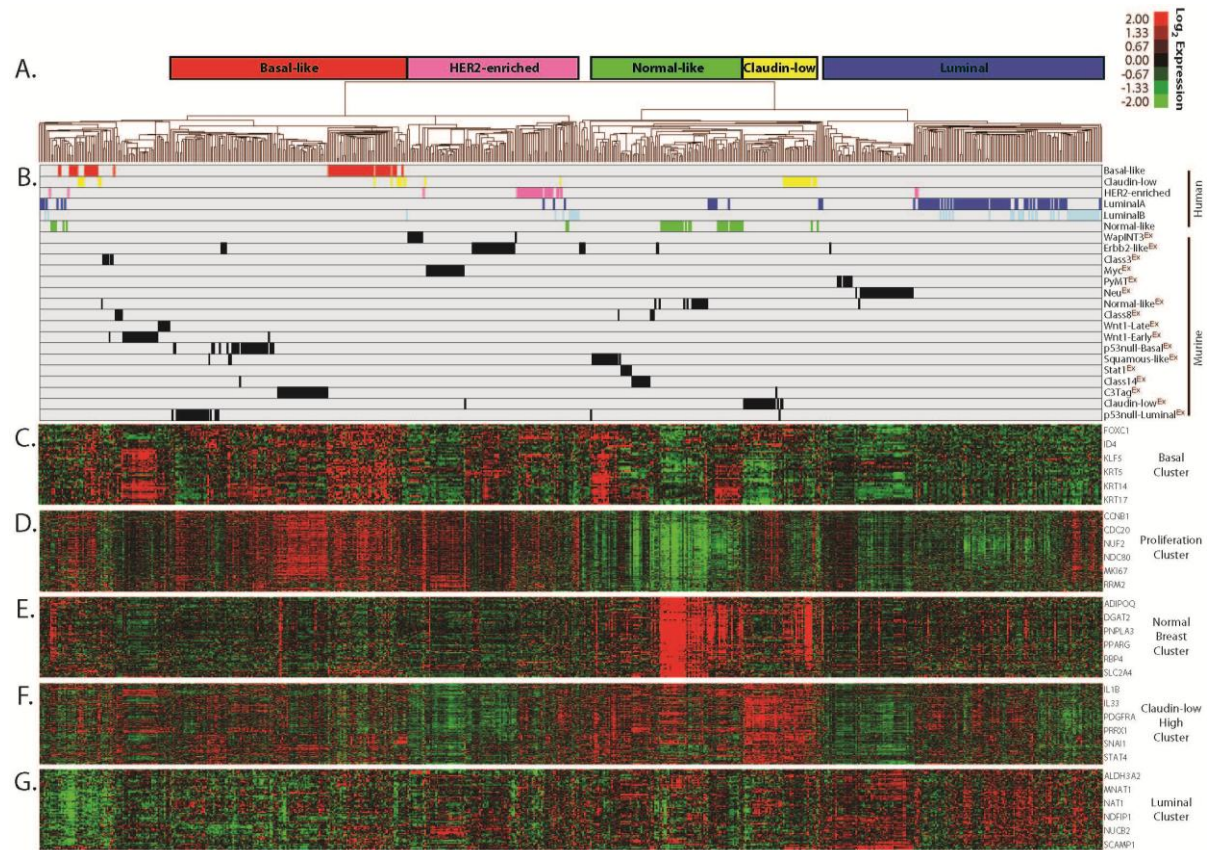


Figure 6: Human and murine intrinsic co-cluster

A. Supervised cluster using a combined human and mouse intrinsic gene list and all murine and UNC308 human arrays. Broad tumor clusters are highlighted with names corresponding to the major human subtype(s) found within each. **B.** Clustering location of all tumors as displayed by their human subtype or mouse class. **C.** Basal gene cluster. **D.** Proliferation gene cluster. **E.** Normal breast gene cluster **F.** Claudin-low subtype high expression gene cluster **G.** Luminal gene cluster.

Human Breast Cancer Subtype																			Predicted Human Counterpart
Mouse Class	Basal-like			Claudin-Low			HER2 Enriched			Luminal A			Luminal B			Normal-Like			
	UNC	p-value Combined	TCGA	UNC	p-value Combined	TCGA	UNC	p-value Combined	TCGA	UNC	p-value Combined	TCGA	UNC	p-value Combined	TCGA	UNC	p-value Combined	TCGA	
WapINT3 ^{Ex}	0.06	0.09	0.17	-	-	NA	-	-	-	-	0.44	-	0.40	0.34	0.29	-	-	-	
Erbb2-like ^{Ex}	0.33	0.30	0.33	-	-	NA	<1e-4	0.01	0.01	0.31	-	-	0.44	0.40	0.30	-	-	-	HER2 Enriched
Class3 ^{Ex}	-	-	-	0.46	-	NA	0.41	0.17	0.38	0.31	0.28	0.34	-	-	-	0.12	0.14	0.29	
Myc ^{Ex}	0.02	0.01	0.03	-	-	NA	0.22	0.11	0.07	-	-	-	0.06	0.01	0.02	-	-	-	Basal-like/LumB
PyMT ^{Ex}	0.41	0.38	-	-	-	NA	0.28	0.09	0.08	0.08	0.33	0.46	0.02	0.10	0.12	-	-	-	
Neu ^{Ex}	-	-	-	-	-	NA	0.44	0.36	0.42	<1e-4	0.01	0.02	0.10	0.36	0.43	-	-	-	Luminal A
Normal-like ^{Ex}	-	-	-	0.14	0.21	NA	-	-	-	-	0.07	0.11	-	-	-	<1e-4	0.01	5e-4	Normal-like
Class8 ^{Ex}	-	-	-	0.09	0.06	NA	0.48	-	-	0.40	0.46	0.11	-	-	-	0.28	0.25	0.26	
Wnt1-Late ^{Ex}	0.37	-	-	-	-	NA	-	-	-	0.40	0.41	0.42	-	0.46	0.40	0.15	0.01	0.21	
Wnt1-Early ^{Ex}	0.29	-	-	-	-	NA	-	-	-	0.40	0.19	0.33	0.38	0.40	0.49	0.39	0.08	0.21	
p53null-Basal ^{Ex}	0.04	0.05	0.06	-	-	NA	-	-	0.16	-	-	-	0.48	0.29	0.20	-	-	-	Basal-like
Squamous-like ^{Ex}	-	-	0.35	0.11	0.02	NA	0.20	-	-	-	-	-	-	-	-	0.18	0.09	0.10	
Stat1 ^{Ex}	-	-	-	0.37	0.32	NA	0.07	-	-	0.31	0.30	0.16	-	0.48	0.41	0.38	0.39	-	
Class14 ^{Ex}	-	-	-	0.35	0.22	NA	-	-	-	0.17	0.14	0.01	0.45	-	0.11	0.06	<1e-4	0.04	Normal-like
C3Tag ^{Ex}	0.02	0.02	0.03	0.38	-	NA	-	-	0.24	-	-	-	0.28	0.12	0.02	-	-	-	Basal-like
Claudin-low ^{Ex}	-	-	0.38	5e-4	<1e-4	NA	-	-	-	-	-	0.20	-	-	0.41	-	-	0.17	Claudin-low
p53null-Luminal ^{Ex}	0.17	0.07	0.02	-	-	NA	0.35	0.23	0.15	-	-	-	0.24	0.24	0.16	-	-	-	

Table 3: Gene set analysis of murine classes and human subtypes

Displayed are the p-values for the gene set analysis comparison of each murine class versus each human subtype. Empty boxes are trending associations, while filled boxes are significant associations (p<0.05, FDR<0.1).

platforms, different common references) may limit interspecies clustering, several across species dendrogram nodes were observed (Figure 6A). Interestingly, all major nodes contained a combination of human and mouse subtypes (Figure 6B), indicating a degree of similarity not only between specific corresponding tumor subtypes, but also globally across species. Most of the major intrinsic gene sets driving the nodes are highlighted below the dendrogram, including the basal (2.4C), proliferation (2.4D), normal breast (2.4E), claudin-low subtype high expression (2.4F), and luminal (2.4G) signatures. These clusters highlight the broad conserved intrinsic features between mouse and human tumors. For instance, most C3Tag^{Ex} tumors cluster with the basal-like subtype, an association that is driven in part by the high expression of the proliferation gene set [31], which is known to contain many E2F-regulated genes.

To more objectively validate the trans-species associations observed in Figure 6, similarity between specific human and mouse subtypes was measured GSA (Table 3) [42]. Using this approach, a murine class was judged to be a strong human subtype counterpart if the human-to-mouse comparison was statistically significant ($p \leq 0.05$) in at least two of the three human datasets analyzed (UNC308 [34], Combined855 [43], and TCGA547 [39]).

As previously observed [31], the murine Normal-like^{Ex}, C3Tag^{Ex}, and Claudin-low^{Ex} classes associate with the human normal-like, basal-like, and claudin-low subtypes, respectively. The new murine class, Erbb2-like^{Ex}, was associated with the human HER2-enriched subtype across all three human data sets; this human breast cancer subtype did not associate with any previously characterized murine class [31], indicating an increased ability for the current dataset to encompass more of the major human intrinsic subtypes. With this larger sample size, a link was also identified between the Myc^{Ex} class and human basal-like

breast cancer, which is consistent with multiple human studies linking basal-like breast cancers with cMYC amplification and expression signatures [39, 44]. Interestingly, a connection between the Myc^{Ex} class and human luminal B tumors was also identified, highlighting Myc activation as a potentially important etiological mechanism that is shared between these two aggressive human subtypes.

Previously defined as a ‘luminal’ model [31], the Neu^{Ex} murine class associated with the human luminal A subtype in this newest analysis; this correlation was somewhat surprising given the lack of ER α and ER α -regulated gene expression in the murine Neu^{Ex} class, but does suggest that human Luminal A tumors have many ER α independent features. Although the murine p53null-Basal^{Ex} versus human comparisons were not significant after controlling for multiple comparisons, an almost consistent significant association was seen with human basal-like tumors (p=0.04, 0.05, and 0.06) in all three human datasets. Lastly, Class14^{Ex} tumors were identified as a counterpart for normal-like human tumors, and of the 13 murine tumors comprising this class, 38% (5/13) are from the Pik3ca-H1047R model. This class clusters independent of normal mammary tissue samples (which are all classified as Normal-like^{Ex}), indicating that this association is possibly not driven by contamination of normal tissue in the tumor biopsies.

Conserved tumorigenic pathway signatures identified between human-mouse counterparts

Many researchers have hypothesized that gene expression signatures may be a more robust means of utilizing gene expression data for discovery and pathway-based classification as they are composed of tens to hundreds of coordinately expressed genes. To

take advantage of this approach, the median expression values for 963 publicly available pathway gene-signatures were calculated separately for the mouse and human datasets, and a two-class (Class X versus all others) Significance Analysis of Microarrays (SAM) was used to identify pathways that were highly expressed by each class/subtype with a false discovery rate (FDR) of 0%. To visualize pathway similarities across species, gene signatures highly expressed within each mouse class were first grouped into ‘pathway meta-signatures’, similar to the way coordinately expressed genes can be grouped into ‘gene signatures’. The average value of these ‘pathway meta-signatures’ was then calculated for each human tumor and displayed as standardized boxplots based on their human breast cancer subtype for the eight mouse classes with human counterparts (Figure 7). These boxplots allow for broad trends to be observed between the pathways highly expressed within each mouse class relative to human tumors, and in all instances, identified tens of pathway signatures that were commonly expressed across species. For instance, the average expression of the 135 pathway signatures highly expressed in C3-Tag^{Ex} tumors were also very highly expressed in human basal-like tumors (Figure 7, top left panel), consistent with the gene level analysis. While these trends are informative, it was of most importance to identify the specific pathways that were highly expressed in both mouse and their human counterparts; it is likely that these shared pathways provide etiological insight and highlight potentially important cancer driving pathways. A subset of the pathways identified as highly expressed in both human and mouse counterparts are displayed below each graph.

Three murine classes overlapped with human basal-like tumors (Figure 7). One common feature between these human and mouse tumors included Trp53 loss/mutation, which in human basal-like tumors occurs in >85% of the samples [39]. This trait was most

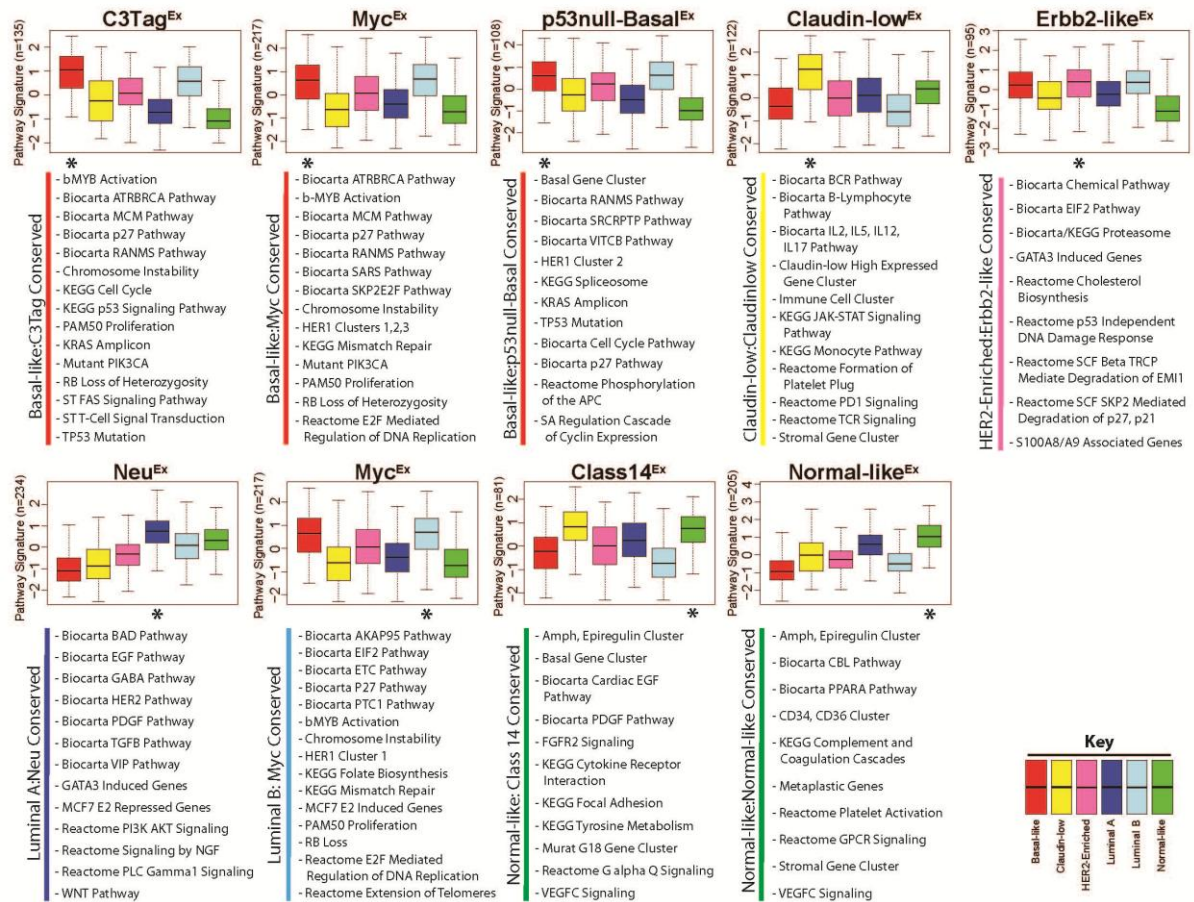


Figure 7: Conserved signaling pathways between human-mouse counterparts

A two-class SAM (Class X versus all others) was used to identify pathways highly expressed in each murine class. Pathways highly expressed with a FDR of 0% were grouped together to define a ‘pathway meta-signature’ for each murine class (with the total number of pathway signatures included shown on the left axis). The standardized, average expression values of each ‘pathway meta-signature’ were calculated in the UNC308, Combined855, and TCGA547 human datasets, which are displayed as boxplots according to their intrinsic human subtype. A subset of the pathways independently identified to be highly expressed in both human-mouse counterparts (as indicated by the ‘*’) for all three human datasets is displayed below each plot.

apparent in C3-Tag^{Ex} and p53null-Basal^{Ex} murine tumors on both a genetic and expression level. The second cardinal feature of human basal-like tumors is high proliferation, primarily resulting from RB loss [39, 40]. Consistent with this finding, all three basal-like mouse classes highly expressed cell cycle and/or RB-pathway related signatures. In addition, C3Tag^{Ex} tumors were enriched for KRAS amplicon genes, b-MYB activation, mutant PIK3CA, and FAS signaling. Murine Myc^{Ex} tumors were also enriched for b-MYB activation and mutant PIK3CA signaling, in addition to a HER1-pathway signature and E2F signaling. Lastly, the p53null-Basal^{Ex} class was enriched for a SRC activation signature, a HER1-pathway signature, and the KRAS amplicon. These findings are relevant since it has been shown that human basal-like tumors also highly express the b-MYB signature [45], are often KRAS [46] and cMYC amplified [39], and show a PIK3CA-activation signature [39, 47]. Thus for human and murine basal-like cancers, both the underlying molecular genetics and their expression profiles are very similar across species.

Human and mouse claudin-low tumors also share many features, including high expression of immune cell associated genes/signatures (e.g. BCR, PD1, and TCR signaling), which is likely due to consistently infiltrating immune cells. Both human HER2-enriched and murine Erbb2-like^{Ex} tumors highly expressed the EIF2 pathway, GATA3 induced genes, and p53 independent DNA damage response genes. Human Luminal A and murine Neu^{Ex} tumors exhibited high expression levels of several tyrosine kinase associated pathway signatures including EGF, HER2, PDGF, TGF α , and PIK3CA signaling. In support of this EGF/HER2 pathway finding, it was recently shown that TgMMTV-Neu tumors therapeutically respond to lapatinib (a dual EGFR and HER2 inhibitor) treatment [48], as would be predicted by the nature of this transgene. In addition to mimicking human basal-like tumors, the murine

Myc^{Ex} class was also a counterpart for the luminal B subtype. Interestingly, many of the same pathways that were common with basal-like tumors are also shared with luminal B tumors, highlighting potentially important etiological events that are shared between these two aggressive intrinsic subtypes; these features include proliferation/RB related pathways, increased chromosome instability, and altered DNA damage repair mechanisms.

DISCUSSION

Human breast cancer is a genetically complex disease consisting of well characterized molecular subtypes [33, 35]. Mouse models can provide an excellent resource to study human disease, but it is essential to ensure the chosen models accurately replicate genetic alterations and overall phenotypes observed in human tumors. Thus, a number of considerations must be kept in mind when designing and/or selecting GEMMs to mimic the human disease state; these features should include intramodel tumor diversity, the degree of genetic similarity, the degree of transcriptomic similarity, and histological similarity (a topic not addressed here). By consolidating mouse models of breast carcinoma into a single dataset, this study was able to investigate the first three of these issues, in which we identified murine models for all of the major human expression subtypes.

To address intramodel tumor diversity, three types of models were identified based on hierarchical clustering analysis: ‘homogenous’, ‘semi-homogeneous’, and ‘heterogeneous’. ‘Homogeneous’ GEMMs were associated with a single murine expression class and were generally created through the expression of oncogenes, possibly relying less on secondary or tertiary mutations that arise during tumor progression. These GEMMs make good experimental models because the phenotypes of individual tumors are consistent and similar.

‘Semi-homogeneous’ models, such as TgMMTV-*Wnt1*, were associated with two murine classes. We hypothesize that unknown secondary events after the initial transgene lesion determine the class fate of these developing tumors. These varying combinations of secondary lesions may cooperate with aberrant Wnt1 signaling to target different mammary cell populations, contributing to model complexity. The last type of model comprises tumors with ‘heterogeneous’ gene expression patterns (i.e. models showing three or more distinct phenotypes). In contrast to ‘homogeneous’ models, the majority of the ‘heterogeneous’ models were based on disrupting the function of tumor suppressor genes. Again, we hypothesize that secondary events after the initial transgene lesion are involved in the class fate determination of these tumors. For example, the *Trp53*^{-/-} model shows specific DNA copy number changes associated with each expression class [32]. From an experimental perspective, special considerations (i.e. phenotyping each individual tumor) must be made to account for this heterogeneity, especially when these models will be utilized for therapeutic efficacy testing.

Despite the diversity of the models tested here, we found that these mouse models collapse into distinct murine classes which recapitulate specific human subtypes on a gene expression-based level. These results are important as they allow for the identification of shared characteristics/lesions between murine and human tumors, and they direct researchers toward appropriate *in vivo* models of specific human subtypes for future experimental testing. Basal-like breast tumors are one the most aggressive subtypes of breast cancer. Herein, we find that three murine classes recapitulated human basal-like breast cancers: C3Tag^{Ex}, Myc^{Ex}, and p53null-Basal^{Ex}. The human basal-like subtype is characterized by high proliferation [49], genomic instability [46], and expression of a c-MYC signature [39, 44].

These murine classes share these hallmarks as evident by high expression of the proliferation gene cluster, cell cycle pathways, and chromosome instability gene-signatures; thus there are clear GEMMs of human basal-like tumors that share both common genetic drivers and expression features.

Murine Claudin-low^{Ex} tumors were identified that significantly mimic the human claudin-low subtype; however, no homogeneous murine model was specific to this class/subtype. Instead, rare tumors from multiple heterogeneous models coalesced into the murine claudin-low group. As an experimental solution to this heterogeneous GEMM complication, the T11 orthotopic, transplantable syngeneic model was derived from a Claudin-low^{Ex} BALB/c *Trp53*^{-/-} tumor (753R), which maintains its claudin-low expression features even after multiple transplant passages [32]. This transplantable model has been used for extensive therapeutic testing [48], thus suggesting that one method of ‘capturing’ a heterogeneous model in a single state can be accomplished via the serial transplantation of a phenotypically characterized individual tumor. As in the human claudin-low subtype, *Trp53* mutation/loss was a common genetic event in mouse Claudin-low^{Ex} tumors. Similarly, both species highly express EMT related genes, inflammatory gene-signatures, and have low expression of many epithelial cell adhesion genes including E-cadherin [34].

Discovered here was the Erbb2-like^{Ex} murine class, which associated with human HER2-enriched tumors even without highly expressing the *ErbB2* gene; no mouse model from our previous studies mimicked this aggressive human tumor subtype. One homogeneous model was found within this class, namely WapCre-*Etv6*. This model expresses the *Etv6-Ntrk3* fusion gene product, a protein that has been associated with

secretory breast cancers [50]. Consistent with this, we observed that murine Erbb2-like^{Ex} tumors highly express a gene signature in common with lactating normal mammary tissue.

For the human luminal breast cancer subtypes, our previous study identified that the TgMMTV-*Neu* model represents the luminal subtypes more than it resembles HER2-enriched tumors [31]. We provide further evidence here that the murine Neu^{Ex} class specifically associates with human luminal A tumors. Conserved with humans, murine Neu^{Ex} tumors highly express several tyrosine kinase pathway related gene-signatures, namely EGFR and HER2, which would be expected based upon the nature of the Neu/ERBB2 transgene. It has been shown that TgMMTV-*Neu* tumors regress with lapatinib treatment [48], giving credence to our approach for identifying drug targetable driver/maintenance pathways in these tumors using a computational pathway-based approach. Interestingly, only the murine Myc^{Ex} class was shown to consistently associate with luminal B tumors. Since the Myc^{Ex} class was also identified as a basal-like model, aberrant Myc activation may be a common hallmark of these two aggressive subtypes.

While our main focus was to identify human-to-mouse disease counterparts, about half of the mouse classes did not statistically associate with specific human subtypes by our broad analysis. Several of these mouse specific classes, however, had clear basal-like tumor expression features including WapINT3^{Ex}, Wnt1-Late^{Ex}, Wnt1-Early^{Ex}, and Squamous-like^{Ex}. Unlike the other three, the Squamous-like^{Ex} class consisted of a variety of models (e.g. *Pik3ca*-H1047R, *Brg1*^{+/-}, and DMBA-induced) and trended toward an association with human claudin-low tumors. Similarly, several classes had luminal expression features, highlighted by PyMT^{Ex} and Stat1^{Ex}. Although the PyMT^{Ex} class had a relatively small number of samples, these tumors trended toward an association with the luminal B subtype. The Stat1^{Ex}

class also had several strong luminal features, consistent with prior characterization of this model [11]. Given the expression of ER α in these *STAT1*-deficient tumors [11], the lack of an association with either the luminal A or luminal B human subtypes was unexpected.

An unanswered question concerning these human-to-mouse associations is the finding that murine classes like Erbb2-like^{Ex}, and Neu^{Ex}, associate with specific human subtypes despite the fact that they apparently do not show expression of one of these human subtype defining genes (*HER2/ERBB2* in the case of Erbb2-like^{Ex} and *ESR1* in the case of Neu^{Ex}). Three hypotheses that could explain this finding are: 1) the cell type of origin of the tumor (but not a genetic driver) is the same across species and this is the major linking phenotype, 2) additional unknown genetic driver(s) are responsible for the common phenotype across species, or 3) some combination of hypothesis 1 and 2. We favor the common cell type of origin hypothesis, but additional experiments like lineage tracing will be required to unequivocally determine this.

Related to this, there are at least two confounding features within our dataset that should also be considered when interpreting these results. First, most of the oncogene-driven mouse models analyzed here used either the MMTV or WAP promoter in their design. If the activity of these promoters varies as a function of specific mammary cell types, such as luminal versus myoepithelial cells, then only those cells that naturally use these promoters would ever give rise to a tumor in these models; we note that most of the MMTV or WAP driven tumors were luminal. Second, similar complications potentially exist with regards to mouse strain. Varying the background genetics in which a model is designed can influence tumor phenotype, and thus classification. Unfortunately, our dataset is underpowered to

adequately address these two confounding features, but future experiments/models could be designed to address these questions.

While some of the mouse classes were identified as good counterparts for specific human subtypes, many were not. There are several possibilities to explain this lack of association. The first is that these classes are specific to murine mammary carcinomas and do not have a matching counterpart in humans. The second might be that these murine classes model rare phenotypes that exist in only a small subset of human breast cancer patients, and that these rare human subtypes were not present in the datasets used here. Similarly, more mouse tumors for classes with small numbers may be required to increase statistical power to detect an association; for example, we hypothesize this to be the case for the PyMT^{Ex} class. The third possibility is that these novel murine classes share phenotypes with multiple human subtypes, and thus may never be classified as being similar to a single human subtype. Some murine tumor features were shared across multiple human subtypes (e.g. Myc^{Ex} with human basal-like and luminal B), which our presented analysis is more likely to undervalue.

While this study provides a framework for identifying GEMMs that could be useful for preclinical drug testing, the simultaneous analysis of 27 mouse models restricted our trans-species comparisons to only expression based analyses. The scope of our future work will focus on using models selected based upon these data for preclinical therapeutic testing to better determine the translational utility of these GEMMs. These experiments are already underway and producing promising results using the TgMMTV-*Neu*, TgC3(1)-*Tag*, and the claudin-low T11 models [48, 51-53]. For example in Roberts *et al.* [51], we showed that the CyclinD1 dependent TgMMTV-*Neu* tumors are sensitive to a CDK4/6 inhibitor, while the

basal-like TgC3(1)-Tag tumor were not; these studies are consistent with findings coming from human clinical trials of luminal/ER+ breast cancers, which were generally noted to be sensitive to a CDK4/6 inhibitor [54]. Similarly, a trans-species genetic screen by Bennett *et al.* [53] identified two ribonucleotide reductase genes (*RRM1* and *RRM2*) and a checkpoint kinase (*CHK1*) as potential targets for triple-negative breast cancer patients, which they validated in both species with drug treatment experiments using TgC3(1)-Tag and human xenograft tumors.

CONCLUSION

In summary, we consolidate 27 murine models of breast carcinoma into the largest comprehensive genomic dataset to date, and we provide a detailed characterization of each to better understand how these GEMMs recapitulate phenotypes of the human subtypes. The data presented here provide insight into the molecular pathways involved in specific breast cancer subtypes and should serve as a useful resource when designing preclinical studies and interpreting their results.

MATERIALS AND METHODS

Gene expression microarrays

A murine tumor dataset of 385 deoxyribonucleic acid (DNA) gene expression microarrays from 27 GEMMs of mammary carcinoma was compiled (Table 1A). 275 of these samples were obtained from multiple previous publications (Gene Expression Omnibus accession numbers: GSE3165, GSE8516, GSE9343, GSE14457, GSE15263, GSE17916, and GSE27101). The other 110 microarray samples (GSE42640) represent newly obtained tumor

samples from multiple participating investigators using methods approved by international animal husbandry guidelines. Total ribonucleic acid (RNA) was purified from 20-30mg of mouse mammary tumor using Qiagen's RNeasy Mini Kit following manufacture protocols. RNA quantity and quality were determined using the Nanodrop spectrophotometer and Agilent Bioanalyzer, respectively. Total RNA was reverse transcribed and labeled with cyanine-5 (Cy5) dye for experimental samples and cyanine-3 (Cy3) dye for mouse reference samples [31] using the Agilent Low RNA Input Fluorescent Linear Amplification Kit. Equal quantities of labeled mouse reference RNA and tumor RNA were co-hybridized overnight to Agilent microarrays, washed, scanned and signal intensities were determined.

All tumor samples were co-hybridized to one of three Agilent Technology gene expression microarray types: 22K, 4X44K, or 4X180K (Figure 1). Two ‘homogeneous expression’ murine models [31], namely *TgMMTV-Neu* and *TgC3(I)-Tag*, were analyzed on all three array types. Therefore, we used both of these models to normalize expression between microarray types [32]. Ten microarrays (five *TgMMTV-Neu* and five *TgC3(I)-Tag*) from each array type were used for normalization (30 microarrays total). All microarray data was independently extracted from the UNC Microarray Database for each array type as \log_2 Cy5/Cy3 ratios, filtering for probes with Lowess normalized intensity values greater than ten in both channels and for probes with data on greater than 70% of the microarrays [31, 34]. Before normalization, each data set was imputed (via the ten-nearest neighbor gene values) and then reduced to the probes that were present on all three array type datasets (11690 probes, 11167 genes). Using the ten normalization arrays per three array platforms, the median expression value was calculated for each probe, on each array type, and a normalization factor was applied independently to each probe so the median was the same for

each array type. Probe expression values were ‘median centered’ to obtain the final normalized dataset. A principle component analysis (PCA) was performed to verify the normalization.

Murine intrinsic genes and subtypes

After removing technical replicates, the dataset was filtered to probes with at least three observations with an absolute \log_2 expression value greater than three using Gene Cluster 3.0 [56], which included 908 probes (899 genes). Hierarchical clustering was performed with this unsupervised probe list using centroid linkage and was viewed with Java Treeview v1.1.5r2 [57]. Potential ‘intrinsic groups’ of murine samples were defined as any set of samples/arrays within this hierarchical cluster that had a Pearson correlation value of 0.65 or greater [31]. Using these defined groups (42 total), an ‘intrinsic gene list’ of 1855 probes (1841 genes) was identified with Intrinsic Gene Identifier v1.0 (Max Diehn/Stanford University) by using a cutoff of one standard deviation below the mean intrinsic gene value [31].

To identify significant murine ‘intrinsic subtypes’, the 385 sample dataset was clustered again using the 1855 intrinsic probe list and SigClust [37] was used to identify groups of samples with a significant association to one another ($p < 0.01$) [32]. GEMM classes were defined as having at least five tumors and a SigClust p-value ≤ 0.01 , yielding 17 classes. Class specific probes/genes were determined using a two class (class X versus all other samples) SAM analysis (v3.11) [34, 58].

Human and mouse intrinsic gene cocluster

Prior to combining the two datasets, probes corresponding to orthologous gene IDs (as determined by the Mouse Genome Informatics of the Jackson Laboratory) were averaged for both the mouse and UNC308 human datasets. Using only orthologous genes found in both datasets (8034 genes), each tumor and gene was standardized to have an average expression of zero and a standard deviation of one ($N(0,1)$) separately for each species. Then, the datasets were merged and each gene was median centered to obtain the final, normalized combined dataset. A merged intrinsic gene list was created by combining the 1841 mouse intrinsic genes defined here and the 1918 human intrinsic genes from Parker *et al* [41] (3310 unique genes in the combined gene set). An intrinsic gene set hierarchical co-cluster was performed using centroid linkage in Gene Cluster 3.0.

Comparison of murine and human expression subtypes

To identify possible commonalities between mouse classes and the human intrinsic subtypes of breast cancer [34, 41], we used the gene set analysis (GSA) R package v1.03 [42] and R v2.12.2. Human subtype specific gene lists were derived for each subtype with a two class (subtype X versus all other samples) SAM analysis independently for all of the unique primary tumor samples from Prat *et al* 2010 (referred to as the UNC308 dataset) [34], from Harrell *et al* 2011 (Combined855 dataset) [43], and from TCGA 2012 (TCGA547 dataset) [39]. Human subtype-specific genes were classified as being highly expressed in the subtype of interest and having a SAM FDR of 0%. Murine classes were then analyzed for significant overlap with each dataset's human subtype-specific gene sets using GSA. Significant overlap was defined as having $p \leq 0.05$ and $FDR \leq 0.1$ to control for multiple comparisons [42].

These same methods were used to identify significant overlap between our 17 newly derived murine classes and the 10 previously defined GEMM classes from Herschkowitz *et al* 2007 [31], noting that all 122 arrays used for the Herschkowitz *et al* study were also present within the 385 sample dataset used here.

Conserved pathway gene signatures

Only genes that were found in both the human and murine datasets were considered for gene-signature analysis in order to eliminate the influence of genes found in only one dataset. Prior to calculating gene-signature values, the human and murine datasets were separately collapsed by averaging rows corresponding to the same gene symbol. Median expression values were calculated for 963 publically available pathway-based gene signatures using methods described in Fan *et al* 2011 [59, 60]. A two class SAM (class or subtype X versus all other samples) was used to identify pathway-signatures enriched in murine and human classes/subtypes, which were defined as being upregulated with a FDR of 0%.

REFERENCES

1. Society AC: **Cancer Facts & Figures 2011.** *Cancer Facts & Figures* 2011.
2. Toft DJ, Cryns VL: **Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies.** *Mol Endocrinol* 2011, **25**:199-211.
3. Schlotter CM, Vogt U, Allgayer H, Brandt B: **Molecular targeted therapies for breast cancer treatment.** *Breast Cancer Res* 2008, **10**:211.
4. Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nat Rev Drug Discov* 2004, **3**:711-715.
5. Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature* 2012, **483**:531-533.
6. Hutchinson L, Kirk R: **High drug attrition rates--where are we going wrong?** *Nat Rev Clin Oncol* 2011, **8**:189-190.
7. 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.
8. Pei XH, Bai F, Smith MD, Usary J, Fan C, Pai SY, Ho IC, Perou CM, Xiong Y: **CDK inhibitor p18(INK4c) is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis.** *Cancer Cell* 2009, **15**:389-401.
9. Adams JR, Xu K, Liu JC, Agamez NM, Loch AJ, Wong RG, Wang W, Wright KL, Lane TF, Zacksenhaus E, Egan SE: **Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation.** *Cancer Res* 2011, **71**:2706-2717.
10. 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.
11. Chan SR, Vermi W, Luo J, Lucini L, Rickert C, Fowler AM, Lonardi S, Arthur C, Young LJ, Levy DE, et al: **STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas.** *Breast Cancer Res* 2012, **14**:R16.
12. Torres-Arzayus MI, Font de Mora J, Yuan J, Vazquez F, Bronson R, Rue M, Sellers WR, Brown M: **High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene.** *Cancer Cell* 2004, **6**:263-274.

13. Liu S, Umez-Goto M, Murph M, Lu Y, Liu W, Zhang F, Yu S, Stephens LC, Cui X, Murrow G, et al: **Expression of autotaxin and lysophosphatidic acid receptors increases mammary tumorigenesis, invasion, and metastases.** *Cancer Cell* 2009, **15**:539-550.
14. Muller WJ, Lee FS, Dickson C, Peters G, Pattengale P, Leder P: **The int-2 gene product acts as an epithelial growth factor in transgenic mice.** *Embo J* 1990, **9**:907-913.
15. 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**:465-475.
16. Pond AC, Herschkowitz JI, Schwertfeger KL, Welm B, Zhang Y, York B, Cardiff RD, Hilsenbeck S, Perou CM, Creighton CJ, et al: **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.
17. Li Z, Tognon CE, Godinho FJ, Yasaitis L, Hock H, Herschkowitz JI, Lannon CL, Cho E, Kim SJ, Bronson RT, et al: **ETV6-NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex.** *Cancer Cell* 2007, **12**:542-558.
18. Cressman VL, Backlund DC, Hicks EM, Gowen LC, Godfrey V, Koller BH: **Mammary tumor formation in p53- and BRCA1-deficient mice.** *Cell Growth Differ* 1999, **10**:1-10.
19. 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.
20. 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.
21. 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.
22. 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.

23. 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.
24. 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.
25. 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.
26. 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.
27. Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD, Van Dyke T: **pRb inactivation in mammary cells reveals common mechanisms for tumor initiation and progression in divergent epithelia.** *PLoS Biol* 2004, **2**:E22.
28. Husler MR, Kotopoulis KA, Sundberg JP, Tennent BJ, Kunig SV, Knowles BB: **Lactation-induced WAP-SV40 Tag transgene expression in C57BL/6J mice leads to mammary carcinoma.** *Transgenic Res* 1998, **7**:253-263.
29. Jerry DJ, Kittrell FS, Kuperwasser C, Laucirica R, Dickinson ES, Bonilla PJ, Butel JS, Medina D: **A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development.** *Oncogene* 2000, **19**:1052-1058.
30. 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.
31. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, et al: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors.** *Genome Biol* 2007, **8**:R76.
32. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, Knezevic J, Greene SB, Darr D, Troester MA, et al: **Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells.** *Proc Natl Acad Sci U S A* 2011, **109**:2778-2783.

33. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
34. 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.
35. Prat A, Perou CM: **Deconstructing the molecular portraits of breast cancer.** *Mol Oncol* 2010.
36. Sharpless NE, Depinho RA: **The mighty mouse: genetically engineered mouse models in cancer drug development.** *Nat Rev Drug Discov* 2006, **5**:741-754.
37. Liu Y, Hayes D, Nobel A, Moarron J: **Statistical Significance of Clustering for High-Dimension, Low-Sample Size Data.** *American Statistical Association* 2008, **103**.
38. Shackleford GM, MacArthur CA, Kwan HC, Varmus HE: **Mouse mammary tumor virus infection accelerates mammary carcinogenesis in Wnt-1 transgenic mice by insertional activation of int-2/Fgf-3 and hst/Fgf-4.** *Proc Natl Acad Sci U S A* 1993, **90**:740-744.
39. Network TCGA: **Comprehensive molecular portraits of human breast tumors.** *Nature* 2012.
40. Herschkowitz JI, He X, Fan C, Perou CM: **The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas.** *Breast Cancer Res* 2008, **10**:R75.
41. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, et al: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**:1160-1167.
42. Efron B, Tibshirani R: **On testing the significance of sets of genes.** *Annals of Applied Statistics* 2007, **1**:107-129.
43. Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, Anders C, Ewend M, Perou CM: **Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse.** *Breast Cancer Res Treat* 2011, **132**:523-535.
44. Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, Cole MD: **A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response.** *PLoS One* 2009, **4**:e6693.

45. Thorner AR, Hoadley KA, Parker JS, Winkel S, Millikan RC, Perou CM: **In vitro and in vivo analysis of B-Myb in basal-like breast cancer.** *Oncogene* 2009, **28**:742-751.
46. Weigman VJ, Chao HH, Shabalín AA, He X, Parker JS, Nordgard SH, Grushko T, Huo D, Nwachukwu C, Nobel A, et al: **Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival.** *Breast Cancer Res Treat* 2011.
47. Hutti JE, Pfefferle AD, Russell SC, Sircar M, Perou CM, Baldwin AS: **Oncogenic PI3K mutations lead to NF-kappaB-dependent cytokine expression following growth factor deprivation.** *Cancer Res* 2012, **72**:3260-3269.
48. Roberts PJ, Usary JE, Darr DB, Dillon PM, Pfefferle AD, Whittle MC, Duncan JS, Johnson SM, Combest AJ, Jin J, et al: **Combined PI3K/mTOR and MEK Inhibition Provides Broad Antitumor Activity in Faithful Murine Cancer Models.** *Clin Cancer Res* 2012.
49. Rakha EA, Reis-Filho JS, Ellis IO: **Basal-like breast cancer: a critical review.** *J Clin Oncol* 2008, **26**:2568-2581.
50. Tognon C, Knezevich SR, Huntsman D, Roskelley CD, Melnyk N, Mathers JA, Becker L, Carneiro F, MacPherson N, Horsman D, et al: **Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma.** *Cancer Cell* 2002, **2**:367-376.
51. Roberts PJ, Bisi JE, Strum JC, Combest AJ, Darr DB, Usary JE, Zamboni WC, Wong KK, Perou CM, Sharpless NE: **Multiple roles of cyclin-dependent kinase 4/6 inhibitors in cancer therapy.** *J Natl Cancer Inst* 2012, **104**:476-487.
52. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L, Karginova O, Jordan J, Combest AJ, Wu M, et al: **Predicting drug responsiveness in humans cancers using genetically engineered mice.** *Clin Cancer Res* 2013, **19**:4889-4899.
53. Bennett CN, Tomlinson CC, Michalowski AM, Chu IM, Luger D, Mittereder LR, Aprelikova O, Shou J, Piwinica-Worms H, Caplen NJ, et al: **Cross-species genomic and functional analyses identify a combination therapy using a CHK1 inhibitor and a ribonucleotide reductase inhibitor to treat triple-negative breast cancer.** *Breast Cancer Res* 2012, **14**:R109.
54. Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO, Ettl J, Patel R, Pinter T, Schmidt M, et al: **Results of a randomized phase 2 study of PD 0332991, a cyclin-dependent kinase (CDK) 4/6 inhibitor, in combination with letrozole vs letrozole alone for first-line treatment of ER+/HER2- advanced breast cancer (BC).** *Cancer Research* 2012, **72**:Abstract nr S1-6.

55. Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, McDonald-Smith GP, Gao H, Hennessy L, et al: **Genomic responses in mouse models poorly mimic human inflammatory diseases.** *Proc Natl Acad Sci U S A* 2013, **110**:3507-3512.
56. de Hoon MJ, Imoto S, Nolan J, Miyano S: **Open source clustering software.** *Bioinformatics* 2004, **20**:1453-1454.
57. Saldanha AJ: **Java Treeview--extensible visualization of microarray data.** *Bioinformatics* 2004, **20**:3246-3248.
58. 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.
59. Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, Perou CM: **Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures.** *BMC Med Genomics* 2011, **4**:3.
60. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**:15545-15550.

CHAPTER 3: LUMINAL PROGENITOR AND FETAL MAMMARY STEM CELL EXPRESSION FEATURES PREDICT BREAST TUMOR RESPONSE TO NEOADJUVANT CHEMOTHERAPY¹

OVERVIEW

Background

Mammary gland morphology and physiology are supported by an underlying cellular differentiation hierarchy. Molecular features associated with particular cell types along this hierarchy may contribute to the biological and clinical heterogeneity observed in human breast carcinomas. Investigating the normal cellular developmental phenotypes in breast tumors may provide new prognostic paradigms, identify new targetable pathways, and explain breast cancer subtype etiology.

Methods

We used transcriptomic profiles coming from fluorescence-activated cell sorted (FACS) normal mammary epithelial cell types from several independent human and murine studies. Using a meta-analysis approach, we derived consensus gene signatures for both species and used these to relate tumors to normal mammary epithelial cell phenotypes. We then compiled a dataset of breast cancer patients treated with neoadjuvant anthracycline and

¹This chapter previously appeared as an article in *Breast Cancer Research and Treatment*. The original citation is as follows: Pfefferle AD *et al*, “Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant chemotherapy”, *Breast Cancer Research and Treatment* 2015

taxane chemotherapy regimens to determine if normal cellular traits predict the likelihood of a pathological complete response (pCR) in a multivariate logistic regression analysis with clinical markers and genomic features such as cell proliferation.

Results

Most human and murine tumor subtypes shared some, but not all, features with a specific FACS purified normal cell type; thus for most tumors a potential distinct cell type of ‘origin’ could be assigned. We found that both human luminal progenitor and mouse fetal mammary stem cell features predicted pCR sensitivity across all breast cancer patients even after controlling for intrinsic subtype, proliferation, and clinical variables.

Conclusions

This work identifies new clinically relevant gene signatures and highlights the value of a developmental biology perspective for uncovering relationships between tumor subtypes and their potential normal cellular counterparts.

BACKGROUND

The mammalian breast is a dynamic organ, with major morphological changes occurring during organogenesis, puberty, pregnancy, lactation and involution [1]. Underlying these mammary gland changes is a complex cell hierarchy that supports these processes [2-4]. The simplest model places the multi-potent mammary stem cell (MaSC) at the base of this hierarchy, having extensive, self-regenerative potential [5]. During mammary development, the MaSC has been proposed to divide asymmetrically to produce basal/myoepithelial cells

as well as luminal progenitors (LumProg), which have more restricted proliferative and differentiation capabilities [5]. LumProg cells are capable of further differentiation into mature luminal (MatureLum) cells, such as Estrogen Receptor (ER)-positive ductal epithelium, which have an even more limited proliferative potential and some of which are terminally differentiated [5].

Breast tumors may originate from several, if not all, of the cell types within this complex mammary hierarchy. These various cellular foundations for tumor initiation may help explain the heterogeneous nature of human breast tumors [6], which consist of multiple histological and genomic subtypes; these genomic groups, which are defined by their gene expression profiles, have become known as the intrinsic subtypes of breast cancer and are referred to as basal-like, claudin-low, HER2-enriched, luminal A, and luminal B [7-10]. A simple etiological explanation for these different subtypes involves a one-to-one relationship between each intrinsic subtype and a distinct cell-type-of-origin that largely maintains its phenotypic identity after oncogenic transformation; however, both normal and neoplastic non-stem cells can acquire stem-like properties, suggesting that the normal cell hierarchy model could also include an element of reversibility [11]. This also raises the possibility that molecular features defining tumor subtypes, may be acquired during tumorigenesis [12].

Genetically engineered mouse models (GEMMs) of breast carcinoma develop heterogeneous tumors [13, 14], but the extent to which they represent human disease is an area of active investigation. We previously showed that murine mammary tumors comprise at least 17 distinct intrinsic subtypes/classes, with eight classes being identified as strong human subtype counterparts by gene expression similarity [14]. As with human breast cancer, the degree to which murine models reflect normal mammary epithelial

subpopulations requires further analysis. Characterization of the cellular features of these murine classes is also needed to better determine their preclinical utility, to shed light on trans-species associations [14], and to help interpret preclinical study observations [15-18].

Several studies have independently profiled fluorescence-activated cell sorted (FACS) purified normal mammary cell types from both human [19-21] and murine [22, 23] mammary tissues. Here, we use a meta-analysis approach to compare the transcriptomic profiles from FACS-enriched mammary cell populations to each other and to primary tumors. These data not only identify a number of clinically relevant biomarkers that may be useful for predicting chemotherapy benefit, but also suggest a cell type of origin for many tumor subtypes.

RESULTS

Comparison of human mammary subpopulation transcriptomic datasets

Several groups have independently obtained transcriptomic profiles of normal human breast cells and compared the genomic biology of these different cell types with human tumors [19-21]. In these studies, normal mammary tissues obtained from female donors were FAC sorted using cell surface markers to enrich for specific mammary subpopulations before microarray analysis (Table 4 and Figure 8). While these initial studies were important, the datasets themselves were relatively small (n=12 for Lim *et al.* 2009, n=72 for Shehata *et al.* 2012, n=18 for Prat *et al.* 2013), and few if any comparisons across studies were performed. Importantly, FACS-based cell fractionation can only enrich for specific subpopulations. Therefore, transcriptomic profiles reflect features of other contaminating cell types to varying degrees. As such, study specific biases may be present in any single dataset; therefore, we

Enriched Population	FACS Markers	Species	Source	Abbreviation	Reference
Stroma	CD49fneg, EpCAMneg	Human	Adult	aStr-Lim09	Lim <i>et al</i> 2009
	CD49fneg, EpCAMneg	Human	Adult	aStr-Shehata	Shehata <i>et al</i> 2012
	CD49fneg, EpCAMneg	Human	Adult	aStr-Prat	Prat <i>et al</i> 2013
Stem Cell	CD49fpos, EpCAMneg	Human	Adult	aMaSC-Lim09	Lim <i>et al</i> 2009
	CD49fpos, EpCAMneg	Human	Adult	aMaSC-Shehata	Shehata <i>et al</i> 2012
	CD49fpos, EpCAMneg	Human	Adult	aMaSC-Prat	Prat <i>et al</i> 2013
Luminal Progenitor	CD49fpos, EpCAMpos	Human	Adult	LumProg-Lim09	Lim <i>et al</i> 2009
	CD49fpos, EpCAMpos	Human	Adult	LumProg-Shehata	Shehata <i>et al</i> 2012
	CD49fpos, EpCAMpos	Human	Adult	LumProg-Prat	Prat <i>et al</i> 2013
Mature Luminal	CD49fneg, EpCAMpos	Human	Adult	MatureLum-Lim09	Lim <i>et al</i> 2009
	CD49fneg, EpCAMpos	Human	Adult	MatureLum-Shehata	Shehata <i>et al</i> 2012
	CD49fneg, EpCAMpos	Human	Adult	MatureLum-Prat	Prat <i>et al</i> 2013

Table 4: Human FACS enriched normal mammary cell subpopulation studies

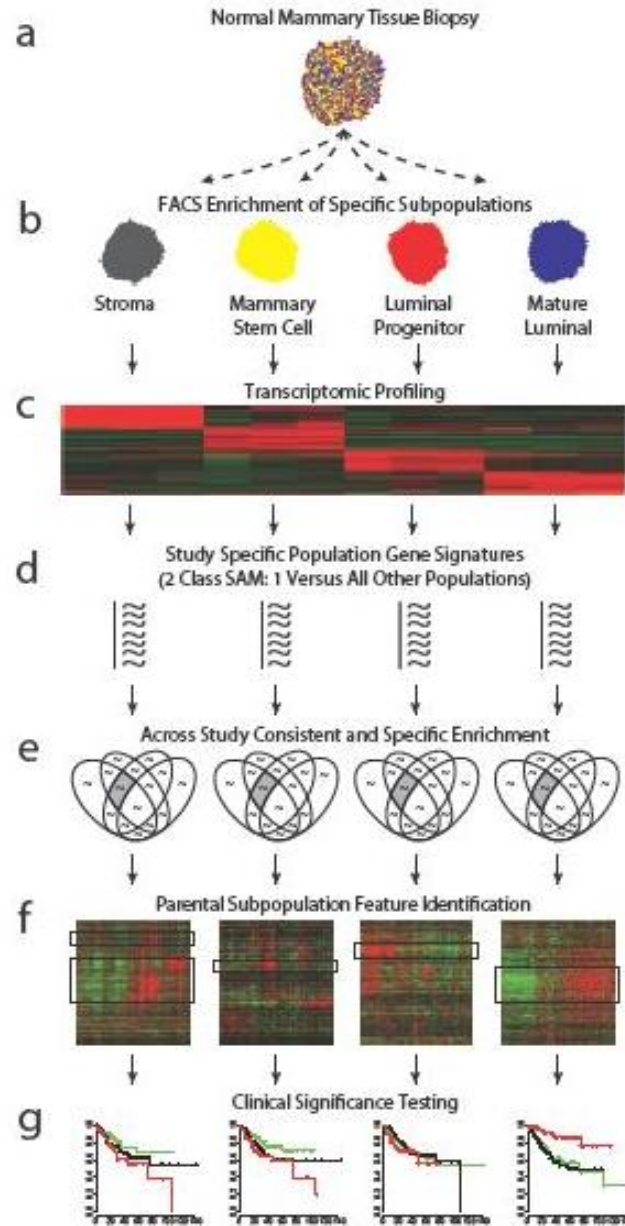


Figure 8: Flowchart of analysis

Normal mammary tissue biopsies were taken from female patients (a) and FACS enriched into distinct mammary cell subpopulations (b). Transcriptome profiling was performed on each subpopulation using gene expression microarrays by three different studies (c). Within each study, genes highly expressed within each subpopulation were determined using a two-class SAM (d). Genes commonly and specifically enriched within each subpopulation across studies were determined to identify ‘enriched’ gene signatures (e). Each ‘enriched’ signature was refined by supervised hierarchical clustering to identify gene ‘features’ highly correlated across a diverse set of human breast tumors (f). These gene signatures were then used for clinical testing (g).

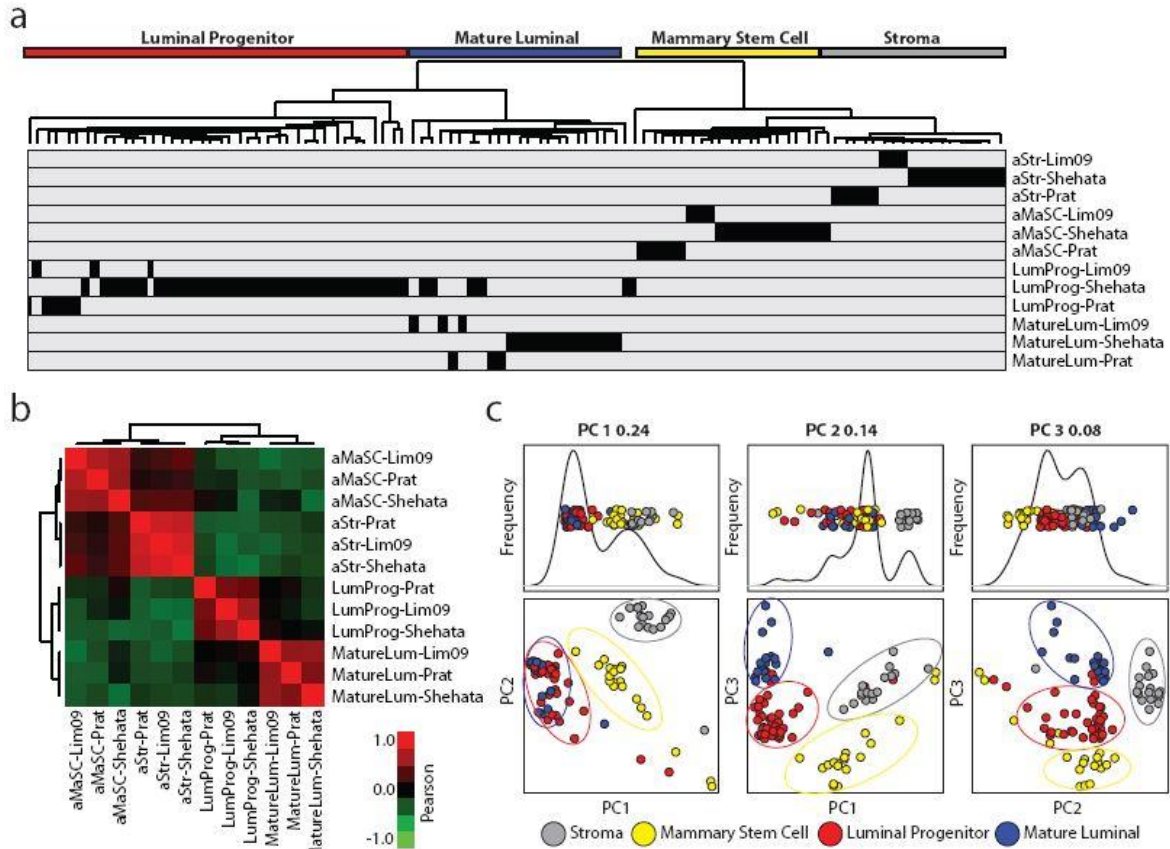


Figure 9: Comparison of mammary subpopulations across studies

a. Unsupervised hierarchical clustering was performed with the normal human mammary subpopulation dataset using any gene that had a \log_2 absolute expression value greater than three in at least four samples. **b.** Pearson correlations were determined between the average expression of each study's subpopulations using all genes. **c.** The first three principle components were determined across the human mammary subpopulation dataset.

used consensus information from all three FACS-enriched human transcriptomic datasets to reduce technical and study-specific biases.

Following DWD normalization [24], an unsupervised cluster of the most variably expressed genes was performed using Gene Cluster v3.0 by selecting all genes with an absolute \log_2 expression value greater than three in at least four samples (212 genes) (Figure 9a). In general, the four major array dendrogram nodes correspond to the four FACS-enriched mammary subpopulations, indicating that the most highly and variably expressed genes are similarly expressed across the different studies. Even when using all genes in the dataset, there is a high Pearson correlation within a given subpopulation across studies and low correlations to other subpopulations (Figure 9b).

On a per-sample basis, the first principle component separated the stroma and adult mammary stem cell (aMaSC) samples from the luminal progenitor and mature luminal samples (Figure 9c). The second principle component separated the stroma and aMaSC samples into distinct groups, while the third principle component separated the luminal progenitor and mature luminal samples into distinct groups. The aMaSC subpopulation displayed the highest level of variation, which is likely attributable to varying degrees of contamination by other cell types.

Human mammary cell subpopulation enriched gene signatures

As shown in Figure 9, there is a natural degree of variation between samples of a given subpopulation. We therefore developed gene signatures for each human mammary subpopulation by integrating consensus information across all three datasets (Table 4) to

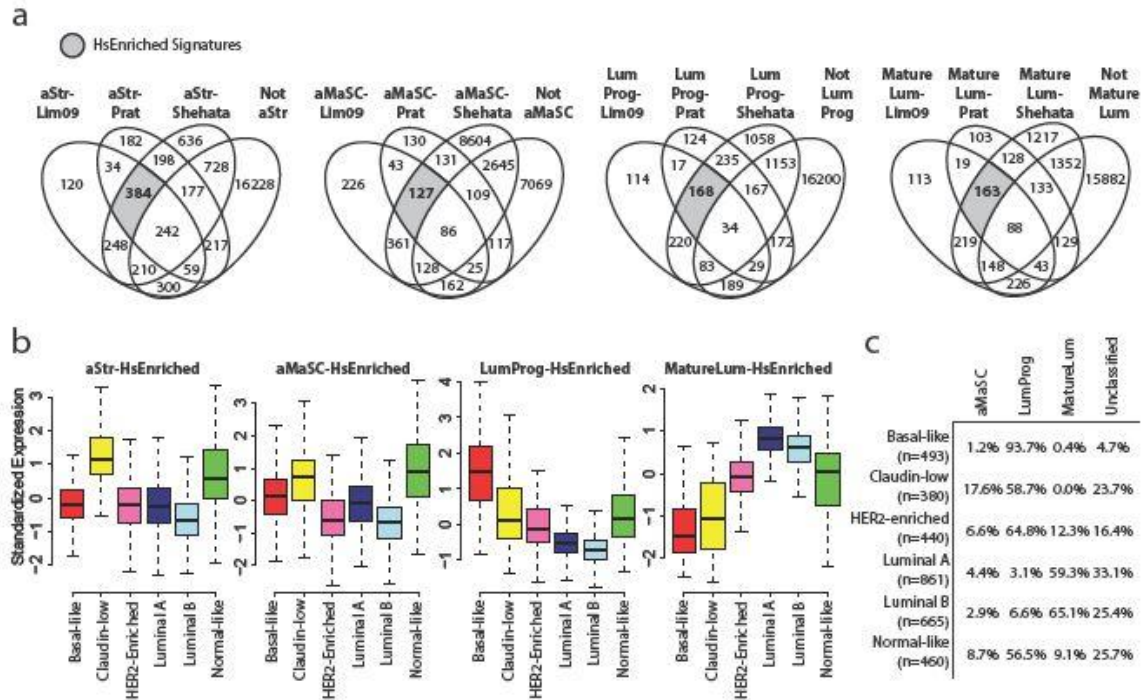


Figure 10: *Homo sapiens* enriched gene signatures

a. HsEnriched gene signatures were identified for each mammary subpopulation. First, the overlap of genes highly expressed within each subpopulation across studies was determined. This overlapping gene set was further filtered to remove genes also identified as enriched in another subpopulation to limit the signature to genes specific to an individual subpopulation. The remaining genes comprised the HsEnriched gene signature for that subpopulation, as indicated by the shaded box. **b.** The standardized average expression of the four HsEnriched gene signatures was calculated across three human datasets and displayed by intrinsic tumor subtype. **c.** A nearest centroid predictor using the HsEnriched gene signatures was used to determine which epithelial features each tumor most represented. To reduce spurious findings, any tumor with a negative silhouette width was considered to have a weak association and was labeled as ‘unclassified’.

identify the highest confidence subpopulation specific genes. First, genes highly expressed ($\text{FDR} < 5\%$) within each mammary subpopulation were found using a two-class (subpopulation X versus all others) SAM analysis [25] within each dataset [19-21]. Second, the overlap of genes highly expressed within a particular subpopulation across studies was determined. Lastly, as it is possible in the above analysis to have the same gene in the signature of more than one subpopulation, genes that were identified to be significantly associated with more than one subpopulation were also removed. This resulted in a single, consensus *Homo sapiens* enriched (HsEnriched) signature per subpopulation (Figure 10a). The average Euclidean distance was determined using a 10-fold cross validation for each normal mammary subpopulation sample to centroids created using either the HsEnriched derived gene signatures or to centroids created using the gene signatures derived separately from each human study. The HsEnriched centroids had a significantly reduced Euclidean distance ($\sim 70\%$) to each mammary subpopulation (t-test $p < 0.0001$), indicating greater specificity for the consensus HsEnriched signatures when compared to any individual dataset's subpopulation signature.

We next evaluated the utility of these signatures for distinguishing human tumor subtypes. Figure 10b displays the standardized average expression of each HsEnriched signature across the human intrinsic breast tumor subtypes [7, 9] using over 3000 tumors [9, 26, 27]. The aStr-HsEnriched signature was highest in claudin-low and normal-like tumors. Interestingly, claudin-low tumors also highly express the aMaSC-HsEnriched signature. High expression of the aMaSC-HsEnriched signature in claudin-low tumors is unlikely an artifact of stromal cells in these tumors since the Pearson correlation between the aStr-HsEnriched and aMaSC-HsEnriched signatures was -0.19 across the normal human mammary samples.

Enriched Population	FACS Markers	Species	Source	Abbreviation	Reference
Stroma	Cd24neg/low/med	Mouse	Fetal	fStr-Spike	Spike <i>et al</i> 2012
	Cd29neg, Cd24neg	Mouse	Adult	aStr-Lim10	Lim <i>et al</i> 2010
Stem Cell	Cd49fhi, Cd24hi	Mouse	Fetal	fMaSC-Spike	Spike <i>et al</i> 2012
	Cd49fhi, Cd24med	Mouse	Adult	aMaSC-Spike	Spike <i>et al</i> 2012
	Cd29pos, Cd24pos, Cd61pos	Mouse	Adult	aMaSC-Lim10	Lim <i>et al</i> 2010
Luminal Progenitor	Cd29neg, Cd24pos, Cd61pos	Mouse	Adult	LumProg-Lim10	Lim <i>et al</i> 2010
Mature Luminal	Cd29neg, Cd24pos, Cd61neg	Mouse	Adult	MatureLum-Lim10	Lim <i>et al</i> 2010

Table 5: Murine FACS enriched normal mammary cell subpopulation studies

The LumProg and MatureLum -HsEnriched signatures were most highly expressed in basal-like and luminal subtype tumors, respectively (Figure 10b).

We noted a considerable degree of signature variation within a subtype, indicating that it is not necessarily the case that all tumors of a given subtype share features with the same normal cell type. A nearest centroid predictor with a 10-fold cross validation error rate of 4.8% was created to individually determine which normal mammary epithelial subpopulation is most similar to each tumor. Samples with positive silhouette widths [28] were considered to have a strong association to their particular subpopulation, with all other tumors being categorized as ‘unclassified’ [29] (Figure 10c). Specifically, 94% of basal-like tumors had LumProg expression profiles. The claudin-low subtype had the highest percentage of tumors classified as aMaSC (18%), although most claudin-low tumors were classified as having LumProg features (59%). The HER2-enriched subtype was predominantly classified as having LumProg expression features. The luminal A and B subtypes were most similar to the MatureLum subpopulation.

Murine mammary cell subpopulation enriched gene signatures

Several groups have also profiled normal murine mammary cell type expression features using FACS [22, 23] (Table 5). In addition to highlighting conserved expression features across species [22], murine studies are uniquely positioned to enable comparisons with developmental states not easily accessed in humans, including early fetal development [23]. We were particularly interested in fetal mammary stem cells (fMaSC) [23], which is a distinct cell population not captured in any human study performed thus far (Table 6). Using the same approach that we used to derive the HsEnriched signatures, we created *Mus*

		Human Populations			
		Str	aMaSC	LumProg	MatureLum
Murine Populations	Str	0.044	—	—	—
	fMaSC	—	—	0.4395	0.4395
	aMaSC	—	0.044	—	—
	LumProg	—	—	0.042	0.386
	MatureLum	—	0.464	0.306	0.004

Table 6: Gene set analysis of human and murine cell subpopulations

A comparative analysis of each human subpopulation versus each murine subpopulation was performed using GSA. The FDR is displayed for all comparisons with a positive association. Statistically significant associations (FDR<0.05) are highlighted.

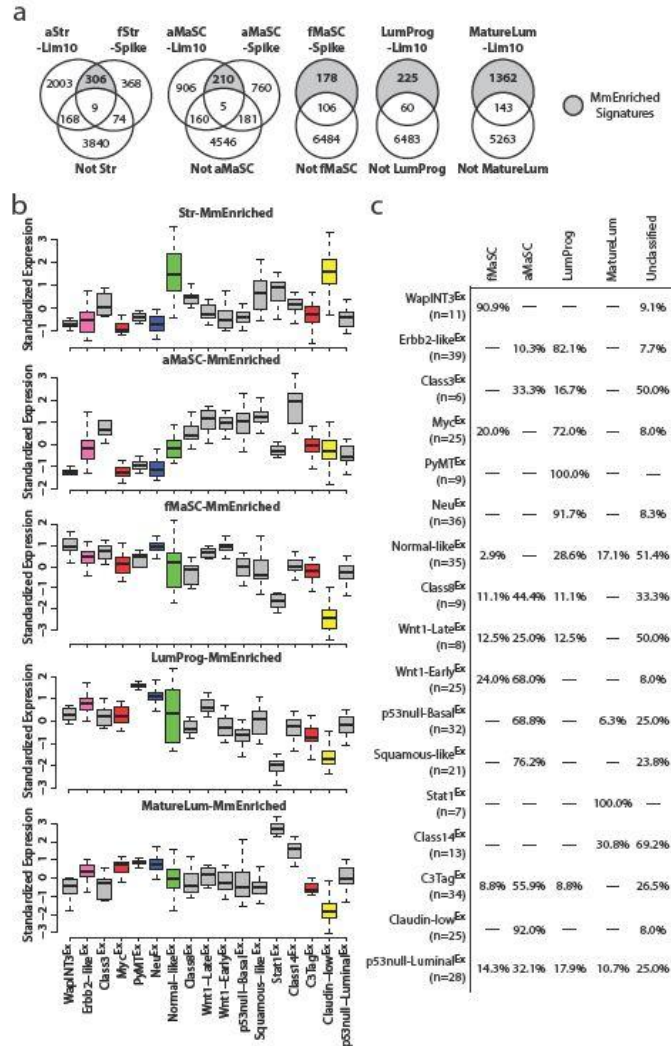


Figure 11: *Mus musculus* enriched gene signatures

a. MmEnriched gene signatures were identified for each mammary subpopulation. First, the overlap of genes highly expressed within each subpopulation across studies was determined. This overlapping gene set was further filtered to remove genes also identified as enriched in another subpopulation to limit the signature to genes specific to an individual subpopulation. The remaining genes comprised the MmEnriched gene signature for that subpopulation, as indicated by the shaded box. **b.** The standardized average expression of the five MmEnriched gene signatures was calculated across a murine dataset and displayed by intrinsic tumor class. **c.** A nearest centroid predictor using the MmEnriched gene signatures was used to determine which epithelial features each tumor most represented. To reduce spurious findings, any tumor with a negative silhouette width was considered to have a weak association and was labeled as ‘unclassified’.

musculus enriched (MmEnriched) signatures for each murine mammary subpopulation (Figure 11a) [22, 23].

We calculated the standardized average expression of each MmEnriched signature across the murine intrinsic subtypes/classes (Figure 11b) [14]. As in human tumors, the Str-MmEnriched signature was most highly expressed in Normal-like^{Ex} and Claudin-low^{Ex}; this common feature was anticipated given the high similarity of these two classes to their human subtype counterparts and their known enrichment for stroma associated genes [14, 23]. The aMaSC-MmEnriched signature was most highly expressed in Class14^{Ex} and to a slightly lesser extent in Wnt1-Late^{Ex}, Wnt1-Early^{Ex}, p53null-Basal^{Ex}, and Squamous-like^{Ex}. The fMaSC-MmEnriched signature was most highly expressed in WapINT3^{Ex}, which is consistent with the finding that *Int3* (*Notch4*) inhibits mammary cell differentiation [30, 31]. The LumProg-MmEnriched signature was highest in PyMT^{Ex} and Neu^{Ex}. This finding was unexpected given that these two mouse classes have been shown to resemble luminal human tumors [13, 14]. Lastly, the MatureLum-MmEnriched signature was most highly expressed in Stat1^{Ex} and Class14^{Ex}. Both the *Stat1*^{-/-} and *Pik3ca*-H1047R mouse models, which define these two classes respectively, are often ER positive [32, 33], and these data suggest that they have mature luminal features. Class14^{Ex} also exhibited significant expression of the aMaSC-MmEnriched signature, indicating that these tumors contain a mixture or share features of multiple cell types.

Consistent with Figure 11b, 91% of WapINT3^{Ex} tumors were classified as having fMaSC features in a nearest centroid predictor analysis. Mouse luminal classes of breast carcinoma (ErbB2-like^{Ex}, Myc^{Ex}, PyMT^{Ex}, and Neu^{Ex}) were most similar to luminal progenitor cells, which again were unexpected but consistent with previous findings [22, 34].

Wnt1-Early^{Ex}, p53null-Basal^{Ex} and Squamous-like^{Ex} tumors had primarily aMaSC features. Interestingly, Claudin-low^{Ex} and to a lesser extent C3-Tag^{Ex} tumors also had aMaSC features. All Stat1^{Ex} tumors had mature luminal features, consistent with being ER positive [32].

LumProg and fMaSC features predict neoadjuvant chemotherapy response

Breast tumors respond heterogeneously to neoadjuvant chemotherapy treatment [15]. We hypothesized that cellular features of normal mammary subpopulations may identify tumors most likely to respond to neoadjuvant chemotherapy. To test this, we compiled a dataset of 702 neoadjuvant anthracycline and taxane chemotherapy treated patients (Table 7).

Although genes within each ‘enriched signature’ are highly correlated within their respective normal cell population, it does not necessarily follow that all genes within a given normal cell signature would be as coordinately regulated in tumors. Therefore, we subdivided each signature into smaller features (feature1, feature2, etc) that are coordinately expressed in tumors, reasoning that such refined ‘features’ may be more clinically robust. All ‘enriched’ and refined ‘features’ were tested for their ability to predict pCR to neoadjuvant chemotherapy in a UVA (Table 8). UVA significant signatures ($p < 0.05$) were then considered in a MVA with Age, ER status, PR status, HER2 status, tumor stage, PAM50 subtype [35], and PAM50 proliferation score [35] to determine if any mammary subpopulation ‘features’ added novel information for predicting pCR (Table 9).

Six normal mammary gene signatures were UVA and MVA significant (Tables 3.5 and 3.6), with the 95% UVA odds ratio of these six signatures and all other ‘enriched signatures’ displayed in Figure 12a. Interestingly, the LumProg-HsEnriched and LumProg-HsEnriched-feature1 signatures, both of which were highly correlated (Figure 12b), were

Characteristic	Total
Age	
< 50	361 (51.4%)
≥ 50	341 (48.6%)
Mean years (SD)	49.7 (10.5)
ER	
negative	311 (44.3%)
positive	391 (55.7%)
PR	
negative	386 (55.0%)
positive	316 (45.0%)
HER2	
negative	654 (93.2%)
positive	48 (6.8%)
Triple Negative	
no	445 (63.4%)
yes	257 (36.6%)
Tumor Stage	
1	33 (4.7%)
2	401 (57.1%)
3	189 (26.9%)
4	79 (11.3%)
pCR	
no	547 (77.9%)
yes	155 (22.1%)
PAM50 Subtype	
Basal-like	233 (33.2%)
HER2-Enriched	67 (9.5%)
Luminal A	208 (29.6%)
Luminal B	141 (20.1%)
Normal-like	53 (7.5%)

Table 7: Clinical characteristics of the neoadjuvant chemotherapy treated dataset

		Patients	UVA p-value	UVA Odds Ratio	UVA AUC
Human FACS Signatures	aStr-HsEnriched	702 (100%)	0.226	0.89 (0.74-1.07)	0.536
	aStr-HsEnriched-feature1	702 (100%)	0.863	0.98 (0.82-1.18)	0.510
	aStr-HsEnriched-feature2	702 (100%)	0.083	0.85 (0.70-1.02)	0.550
	aMaSC-HsEnriched	702 (100%)	0.703	1.04 (0.87-1.24)	0.510
	aMaSC-HsEnriched-feature1	702 (100%)	0.547	0.95 (0.79-1.13)	0.518
	LumProg-HsEnriched	702 (100%)	<0.001	2.12 (1.76-2.58)	0.707
	LumProg-HsEnriched-feature1	702 (100%)	<0.001	1.85 (1.53-2.24)	0.669
	MatureLum-HsEnriched	702 (100%)	<0.001	0.47 (0.38-0.57)	0.704
	MatureLum-HsEnriched-feature1	702 (100%)	<0.001	0.46 (0.37-0.55)	0.713
	Str-MmEnriched	702 (100%)	0.011	0.78 (0.64-0.94)	0.571
Mouse FACS Signatures	Str-MmEnriched-feature1	702 (100%)	0.209	0.89 (0.73-1.06)	0.545
	Str-MmEnriched-feature2	702 (100%)	0.029	0.81 (0.67-0.98)	0.556
	Str-MmEnriched-feature3	702 (100%)	<0.001	0.57 (0.47-0.70)	0.652
	Str-MmEnriched-feature4	702 (100%)	0.008	0.78 (0.65-0.94)	0.574
	fMaSC-MmEnriched	702 (100%)	0.009	1.27 (1.06-1.51)	0.561
	fMaSC-MmEnriched-feature1	702 (100%)	<0.001	1.89 (1.58-2.28)	0.678
	fMaSC-MmEnriched-feature2	702 (100%)	<0.001	0.48 (0.39-0.58)	0.704
	aMaSC-MmEnriched	702 (100%)	0.843	1.02 (0.85-1.22)	0.506
	aMaSC-MmEnriched-feature1	702 (100%)	0.003	1.30 (1.09-1.55)	0.569
	aMaSC-MmEnriched-feature2	702 (100%)	0.491	1.06 (0.89-1.27)	0.530
	aMaSC-MmEnriched-feature3	702 (100%)	0.023	0.81 (0.67-0.97)	0.550
	LumProg-MmEnriched	702 (100%)	0.733	1.03 (0.86-1.23)	0.498
	LumProg-MmEnriched-feature1	702 (100%)	0.488	1.07 (0.89-1.27)	0.527
	LumProg-MmEnriched-feature2	702 (100%)	<0.001	0.56 (0.46-0.68)	0.664
	MatureLum-MmEnriched	702 (100%)	0.004	0.77 (0.64-0.92)	0.583
	MatureLum-MmEnriched-feature1	702 (100%)	<0.001	1.71 (1.42-2.07)	0.653
	MatureLum-MmEnriched-feature2	702 (100%)	<0.001	2.30 (1.89-2.81)	0.729
	MatureLum-MmEnriched-feature3	702 (100%)	0.020	1.25 (1.04-1.50)	0.566
	MatureLum-MmEnriched-feature4	702 (100%)	0.059	0.84 (0.69-1.00)	0.542
	MatureLum-MmEnriched-feature5	702 (100%)	0.164	0.88 (0.74-1.05)	0.548
	MatureLum-MmEnriched-feature6	702 (100%)	0.618	1.05 (0.88-1.25)	0.507
	MatureLum-MmEnriched-feature7	702 (100%)	0.001	1.35 (1.13-1.63)	0.578
	MatureLum-MmEnriched-feature8	702 (100%)	<0.001	0.49 (0.40-0.60)	0.691
	MatureLum-MmEnriched-feature9	702 (100%)	0.291	1.10 (0.92-1.33)	0.522
	MatureLum-MmEnriched-feature10	702 (100%)	0.689	1.04 (0.87-1.24)	0.520
Published Signatures	Usary et al (PMID:23780888)				
	Usary Untreated Hum	702 (100%)	<0.001	2.46 (1.99-3.09)	0.728
	Usary Resp Hum	702 (100%)	0.005	0.76 (0.63-0.92)	0.585
	Hatzi et al (PMID:21558518)				
	Chemosensitive				
	Insensitive	315 (44.9%)		1	
	Sensitive	151 (21.5%)	<0.001	2.58 (1.62-4.10)	0.610
	Not available	236 (33.6%)			
	Chemoresistance				
	RCB_0_I	211 (30.1%)		1	
	RCB_II_III	255 (36.3%)	<0.001	0.10 (0.06-0.18)	0.743
	Not available	236 (33.6%)			
	SET Index				
	Low	411 (58.5%)		1	
	Int	36 (5.1%)	0.070	0.33 (0.08-0.94)	
	High	19 (2.7%)	0.259	0.43 (0.07-1.52)	0.541
	Not available	236 (33.6%)			

Table 8: Univariate logistic regression analysis predicting pathological complete response in breast cancer patients treated with neoadjuvant anthracycline and taxane chemotherapy regimens

		Patients	MVA p-value	MVA Odds Ratio	MVA AUC
Human Signatures	Clinical* + LumProg-HsEnriched	702 (100%)	0.014	1.50 (1.09-2.09)	0.778
	Clinical* + LumProg-HsEnriched-feature1	702 (100%)	0.030	1.39 (1.04-1.88)	0.777
	Clinical* + MatureLum-HsEnriched	702 (100%)	0.433	0.84 (0.54-1.30)	0.772
	Clinical* + MatureLum-HsEnriched-feature1	702 (100%)	0.523	0.86 (0.53-1.38)	0.772
Mouse Signatures	Clinical* + Str-MmEnriched	702 (100%)	0.889	1.02 (0.79-1.31)	0.772
	Clinical* + Str-MmEnriched-feature2	702 (100%)	0.198	0.86 (0.69-1.08)	0.773
	Clinical* + Str-MmEnriched-feature3	702 (100%)	0.048	0.79 (0.62-1.00)	0.776
	Clinical* + Str-MmEnriched-feature4	702 (100%)	0.389	0.91 (0.72-1.13)	0.771
	Clinical* + fMaSC-MmEnriched	702 (100%)	0.216	1.14 (0.93-1.39)	0.773
	Clinical* + fMaSC-MmEnriched-feature1	702 (100%)	0.013	1.48 (1.09-2.03)	0.777
	Clinical* + fMaSC-MmEnriched-feature2	702 (100%)	0.031	0.71 (0.52-0.97)	0.777
	aMaSC-MmEnriched				
	Clinical* + aMaSC-MmEnriched-feature1	702 (100%)	0.749	1.04 (0.82-1.32)	0.772
	Clinical* + aMaSC-MmEnriched-feature3	702 (100%)	0.018	0.77 (0.61-0.95)	0.775
	LumProg-MmEnriched				
	Clinical* + LumProg-MmEnriched-feature2	702 (100%)	0.805	0.96 (0.70-1.33)	0.772
	MatureLum-MmEnriched	702 (100%)	0.697	0.95 (0.73-1.24)	0.772
	Clinical* + MatureLum-MmEnriched-feature1	702 (100%)	0.671	1.06 (0.82-1.37)	0.772
	Clinical* + MatureLum-MmEnriched-feature2	702 (100%)	0.128	1.38 (0.91-2.11)	0.773
	Clinical* + MatureLum-MmEnriched-feature3	702 (100%)	0.461	1.09 (0.87-1.37)	0.772
	Clinical* + MatureLum-MmEnriched-feature7	702 (100%)	0.577	1.07 (0.85-1.34)	0.772
	Clinical* + MatureLum-MmEnriched-feature8	702 (100%)	0.388	0.85 (0.58-1.24)	0.772
Published Signatures	Usary et al (PMID:23780888)				
	Clinical* + Usary Untreated Hum	702 (100%)	0.001	1.90 (1.29-2.83)	0.777
	Clinical* + Usary Resp Hum	702 (100%)	0.021	1.35 (1.05-1.74)	0.774
	Hatzl et al (PMID:21558518)				
	Clinical* + Chemosensitive				
	Insensitive	315 (44.9%)		1	0.817
	Sensitive	151 (21.5%)	<0.001	4.38 (2.50-7.84)	
	Not available	236 (33.6%)			
	Clinical* + Chemoresistance				
	RCB_o_I	211 (30.1%)		1	0.860
	RCB_II_III	255 (36.3%)	<0.001	0.11 (0.06-0.21)	
	Not available	236 (33.6%)			

*Clinical variables consist of Age, ER, PR, HER2, Tumor Stage, PAM50 Subtype, and PAM50 Proliferation Score

Table 9: Multivariate logistic regression analysis predicting pathological complete response in breast cancer patients treated with neoadjuvant anthracycline and taxane chemotherapy regimens

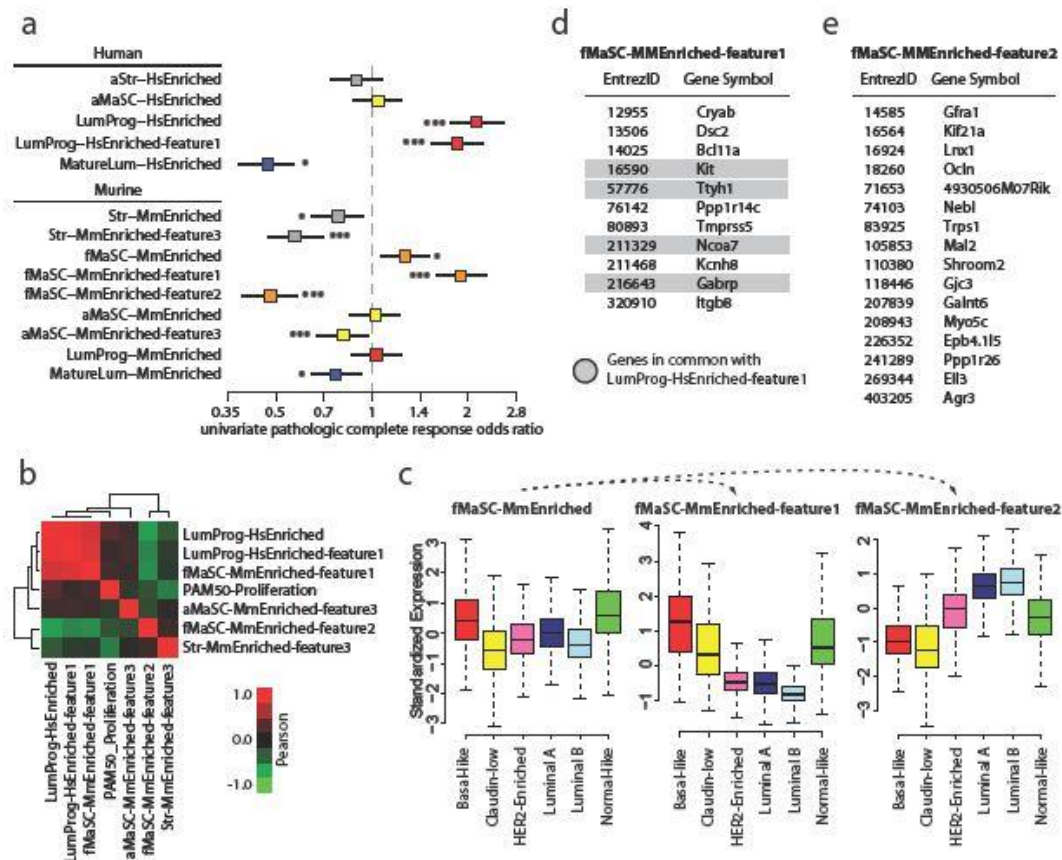


Figure 12: fMaSC enriched gene signatures

a. The univariate logistic regression odds ratio predicting pathologic complete response to neoadjuvant anthracycline and taxane chemotherapy was determined using a 702 patient dataset, with the 95% confidence interval shown as a forest plot. A single ‘*’ indicates the signature was univariate significant, while ‘***’ indicates the signature was both univariate and multivariate significant ($p < 0.05$). **b.** Pearson correlations of multivariate significant gene signatures and proliferation were determined. **c.** The standardized average expression of the fMaSC-MmEnriched signature and its two refined signatures was calculated across three human datasets and displayed by intrinsic tumor subtype. **d.** Genes in the fMaSC-MmEnriched-refined1 signature. **e.** Genes in the fMaSC-MmEnriched-refined2 signature.

highly significant in the UVA and MVA analyses, indicating that tumors with luminal progenitor features are more likely to respond to neoadjuvant treatment. Importantly, this response was independent of proliferation, as highlighted by their low correlation to the PAM50-Proliferation gene signature (Figure 12b).

Interestingly, the fMaSC-MmEnriched signature refined into two distinctly opposite, highly significant signatures in both the UVA and MVA (Table 8, 3.6; Figure 12b, 3.5c). While the fMaSC-MmEnriched signature was highest in basal-like tumors, the refined-signatures varied, with fMaSC-MmEnriched-feature1 (Figure 12d) being highest in basal-like tumors and fMaSC-MmEnriched-feature2 (Figure 12e) expressed in luminal tumors. Tumors with fMaSC-MmEnriched-feature1 expression were more likely to respond to neoadjuvant chemotherapy, while those tumors with fMaSC-MmEnriched-feature2 were more resistant. The fMaSC-MmEnriched-feature1 signature was very highly correlated with the LumProg-HsEnriched signatures (Figure 12b), sharing four genes in common (Figure 12d). These results support the hypothesis that subsets of genes within the larger ‘enriched signature’ are likely regulated by different biological mechanisms.

DISCUSSION

Normal mammary gland physiology is supported by an underlying, complex cell hierarchy [2-5]. The simplest model treats differentiation from mammary stem cells to progenitor cells to mature cells as unidirectional, but recent observations indicate that bidirectional processes are also possible for normal and neoplastic cells [11]. This differentiation plasticity may allow tumors to acquire cell features foreign to the initial cell-

of-origin or to lose native features through the accumulation of specific genetic aberrations [36].

Regardless of how different cellular traits are acquired, it is critical to identify the ‘current’ normal cellular features within a tumor, and therefore, we first analyzed the expression profiles of normal human and mouse mammary epithelial cell subpopulations [19-23]. We chose to use nomenclature that maintains continuity with the literature. However, these terms should be considered provisional as the complete biological profiles of these FACS fractions are investigated [4]. Recent work by Prater *et al* [37] found that mouse ‘luminal progenitor’ cells (CD49f+, EpCAM+) have complete mammary gland repopulating potential, indicating that ‘luminal progenitor’ may be a misnomer. Importantly, even if our understanding and naming of these cell subpopulations change, only the retrospective interpretation of the data presented here will be affected, not the data themselves.

Using a meta-analysis approach, FACS purified mammary epithelial cell subpopulation ‘enriched’ gene signatures were derived and a nearest centroid predictor was developed to identify which normal mammary subpopulation each human and mouse tumor most represented using over three thousand human patients and 27 mouse models of mammary carcinoma [14]. While these analyses imply a cell-of-origin for a given tumor, additional experiments (e.g. lineage tracing) will be required to unequivocally determine this. Nevertheless, these associations at the very least identify which normal mammary subpopulation a given tumor most represents in its current state.

With this in mind, several associations between both the human and mouse intrinsic subtypes and specific normal cell subpopulations were observed. First, human basal-like tumors have been referred to as ‘undifferentiated’, which is consistent with their exhibiting

luminal progenitor [19] and fetal MaSC features [23]. Three mouse classes have been identified to be human basal-like counterparts: Myc^{Ex}, p53null-Basal^{Ex}, and C3-Tag^{Ex} [14]. Myc^{Ex} tumors were the most similar to the luminal progenitor cell profile. By contrast, both p53null-Basal^{Ex} and C3-Tag^{Ex} tumors had adult MaSC features. These results indicate that Myc^{Ex} tumors share similar cell features as their human basal-like counterpart, making it an attractive mouse model for studying basal-like tumors with aberrant Myc signaling [10, 38]. Interestingly, neither p53null-Basal^{Ex} nor C3-Tag^{Ex} tumors had strong luminal progenitor features, indicating that their association to human basal-like tumors is more likely driven by their underlying genetics [10].

Human claudin-low tumors had heterogeneous normal cell features. While most were similar to luminal progenitors, the claudin-low subtype also had the largest percentage of tumors classified as adult MaSC. Given that claudin-low tumors are enriched with epithelial-to-mesenchymal transition features [9, 39, 40], our results suggest that these tumors may originate from the luminal progenitor population prior to acquiring adult MaSC and/or mesenchymal features. Similarly, mouse Claudin-low^{Ex} tumors were also strongly associated with the adult MaSC population, indicating that such tumors may be the closest analogs of the subset of human claudin-low tumors with adult MaSC features.

Human HER2-enriched tumors were the most similar to the luminal progenitor subpopulation. This is a novel finding and may explain why both human basal-like and HER2-enriched subtype tumors show high *TP53* mutation frequencies (>70%) and widespread chromosomal instability [10]. These data could suggest that the normal luminal progenitor cell is somehow extremely dependent upon TP53 function. The murine Erbb2-

like^{Ex} class has been identified as a mouse counterpart for human HER2-enriched tumors [14] and was shown here to also have luminal progenitor features.

When analyzing the human luminal A and B subtypes, a clear association with normal mature luminal cells was observed. The murine Neu^{Ex} class is a proposed counterpart for human luminal A tumors [14], yet these mouse tumors were most similar to normal mouse luminal progenitor cells. The Myc^{Ex} class was also identified to resemble human luminal B tumors [14]. As discussed, Myc^{Ex} tumors have luminal progenitor features; therefore, most mouse luminal A/B tumor models do not share the same normal cell features as their human tumor counterparts. These differences may reflect limitations of model system design, as tumors within these mouse classes are primarily driven by either the WAP or MMTV promoter. These differences in cell features, however, indicate that the trans-species associations observed previously [14] are possibly driven by the genetics of each mouse model. Nevertheless, broad molecular features are conserved between these human-murine counterparts [14]. Therefore, we propose that these mouse models retain significant preclinical utility provided that shared versus distinct molecular features are taken into account.

Neoadjuvant chemotherapy is a common approach for treating breast tumors, but only a relatively low percentage of patients have a pCR (~20% overall). We tested the clinical significance of normal cellular features for predicting pCR using a combination of univariate and multivariate logistic regression analyses. Human LumProg and mouse fetal MaSC expression features were identified as predictive of pCR sensitivity across all breast cancer patients. More specifically, LumProg-HsEnriched-feature1 and fMaSC-MmEnriched-feature1 expression was highest in basal-like tumors. This is consistent with the clinical

observation that basal-like tumors have better neoadjuvant chemotherapy response rates since higher expression of these normal cell signatures was associated with a higher likelihood of pCR. Distinct from these signatures, tumors with high expression of fMaSC-MmEnriched-feature2 were more resistant to neoadjuvant chemotherapy. Not surprisingly, this signature was most highly expressed in luminal A and B tumors, consistent with the clinical observation that these subtypes have lower chemotherapy response rates. Importantly, these signatures remained significant even after controlling for intrinsic subtype, proliferation, and clinical variables in the multivariate analysis; thus these normal cell signatures add information even when tumor subtype and clinical features are known. It is presently unknown whether tumors with these features arise from a LumProg or fetal MaSC cell-of-origin or acquire these features during tumorigenesis. Whether these features are acquired or inherent, the ‘current’ cellular traits of a tumor are likely most important as these appear to be a major determinant of chemotherapy sensitivity. The biological explanation for why LumProg and fetal MaSC expression features predict tumor responsiveness to neoadjuvant chemotherapy will need to be explored further, but it is likely linked to the common genetic features of TP53 loss [41], RB-pathway loss [42], and high proliferation status [43], as well as other inherent characteristics of these cellular states. This work highlights the efficacy of studying the normal mammary gland cell hierarchy and development in providing insights for human tumor therapy responsiveness.

MATERIALS AND METHODS

Mammary cell subpopulation gene signatures

Gene expression measurements from fluorescence-activated cell sorting (FACS) enriched mammary cell subpopulations were obtained from three human and two murine published studies: GSE16997 [19], GSE19446 [22], GSE27027 [23], GSE35399 [20], and GSE50470 [21]. The human and murine datasets were separately combined using distance weighted discrimination (DWD) normalization to adjust for systemic microarray data biases between studies [24]. FACS subpopulation gene signatures were then derived within the human and murine dataset separately using a common approach. First, genes highly expressed within each FACS subpopulation were identified using a two-class (subpopulation X versus all others) Significance Analysis of Microarrays (SAM) analysis [9, 25], with genes highly expressed and with a false discovery rate (FDR) of <5% being considered significant. Next, the intersection of each study's subpopulation gene signature was identified (e.g. $\text{aMaSC-Lim09} \cap \text{aMaSC-Shehata} \cap \text{aMaSC-Prat}$). The intersecting gene set for each cell type was then further limited to genes uniquely found in the subpopulation of interest by removing genes found in any other subpopulation's gene set (e.g. removing members of $\text{aStr-Lim09} \cup \text{aStr-Shehata} \cup \text{aStr-Prat} \cup \text{LumProg-Lim09} \cup \text{LumProg-Shehata} \cup \text{LumProg-Prat} \cup \text{MatureLum-Lim09} \cup \text{MatureLum-Shehata} \cup \text{MatureLum-Prat}$ from the aMaSC intersecting gene set) and by removing genes associated with the myoepithelial subpopulation using a published myoepithelial gene signature produced using the same approach as those derived here [44]. Through this process, a consensus gene signature was produced for each mammary cell FACS subpopulation, for each species, which we designated as 'enriched' (e.g. aMaSC-HsEnriched).

Each FACS ‘enriched’ signature was further refined by supervised clustering using the human UNC308 breast tumor dataset to identify subpopulation ‘features’ [9]. The purpose of this process was to identify clusters of genes highly correlated across a diverse human tumor dataset, as these gene features are more likely regulated by similar factors and therefore, may be more clinically useful than the entire enriched signature. These refined features (e.g. fMaSC-feature1 for example) were defined as having at least ten genes with a Pearson correlation greater than 0.5 across all tumors in the UNC308 dataset [45]. Expression scores for both the ‘enriched’ and ‘feature’ gene signatures were determined by calculating the mean expression of the signature within each tumor. Signatures were separately standardized to have an average expression value of zero and a standard deviation of one ($N(0,1)$) to allow for across signature comparisons.

Comparison of human and murine normal mammary populations

To identify possible commonalities between human and mouse normal mammary FACS populations, we used the gene set analysis (GSA) R package v1.03 [46] and R v2.12.2. Murine populations were analyzed for significant overlap with each HsEnriched gene signature. Significant overlap was defined as having $p \leq 0.05$ and $FDR \leq 0.1$ to control for multiple comparisons [46].

Mammary cell subpopulation centroids

Human mammary cell subpopulation centroids were created using the union of the ‘enriched’ epithelial gene signatures (aMaSC-HsEnriched \cup LumProg-HsEnriched \cup MatureLum-HsEnriched). The DWD single sample predictor (SSP) function [24] was used to

calculate the shortest Euclidean distance between each tumor and each epithelial cell enriched centroid using three human datasets comprising over 3000 patients: UNC308 [9], Combined855 [26], and Metabric2136 [27]. To gauge the strength of each mammary subpopulation association, the silhouette width was calculated for each sample using R v3.0.1 and the ‘cluster’ package. Samples with a positive silhouette width were considered to have strong association. Similarly, this process was repeated using the murine cell subpopulation dataset to calculate Euclidean distances for a murine expression dataset comprising 27 models of mammary carcinoma and normal mammary tissue [14].

Chemotherapy response

Logistic regression analysis was used to determine if gene signatures derived from normal cell populations were capable of predicting pathological complete response (pCR) in breast cancer patients treated with neoadjuvant anthracycline and taxane chemotherapy regimens. For this purpose, a combined breast cancer gene expression dataset was created from three public datasets (GSE25066 [47], GSE32646 [48], and GSE41998 [49]). Only neoadjuvant anthracycline and taxane treated patients with complete clinical data (Age, ER status, PR status, HER2 status, tumor stage and pCR) were considered in the analysis, resulting in a dataset of 702 patients. The three datasets were combined using DWD normalization to adjust for systemic microarray data biases between studies [24], with the clinical characteristics found in Table 7. The significance of each mammary subpopulation gene signature and several published predictors of pCR was determined using a series of stepwise tests. First, the ability for each signature to predict pCR was determined with a univariate analysis (UVA) using R v3.0.1 (Table 8). Those signatures that were significant

($p < 0.05$) were then considered in a multivariate analysis (MVA) with several clinical variables (Age, ER status, PR status, HER2 status, tumor stage, PAM50 subtype [35], and PAM50 proliferation score [35]) to determine if each mammary subpopulation gene signature added new information for predicting pCR (Table 9).

REFERENCES

1. Gjorevski N, Nelson CM: **Integrated morphodynamic signalling of the mammary gland.** *Nat Rev Mol Cell Biol* 2011, **12**:581-593.
2. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C: **Distinct stem cells contribute to mammary gland development and maintenance.** *Nature* 2011, **479**:189-193.
3. Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, Harrell JC, McNamara G, Schwede M, Culhane AC, et al: **Taxonomy of breast cancer based on normal cell phenotype predicts outcome.** *J Clin Invest* 2014, **124**:859-870.
4. Visvader JE, Stingl J: **Mammary stem cells and the differentiation hierarchy: current status and perspectives.** *Genes Dev* 2014, **28**:1143-1158.
5. Visvader JE: **Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis.** *Genes Dev* 2009, **23**:2563-2577.
6. Visvader JE: **Cells of origin in cancer.** *Nature* 2011, **469**:314-322.
7. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
8. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *PNAS* 2001, **98**:10869-10874.
9. 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.
10. TCGA: **Comprehensive molecular portraits of human breast tumours.** *Nature* 2012, **490**:61-70.
11. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, Brooks M, Reinhardt F, Su Y, Polyak K, et al: **Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state.** *PNAS* 2011, **108**:7950-7955.
12. Spike BT, Wahl GM: **p53, Stem Cells, and Reprogramming: Tumor Suppression beyond Guarding the Genome.** *Genes Cancer* 2011, **2**:404-419.
13. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, et al: **Identification of conserved gene**

- expression features between murine mammary carcinoma models and human breast tumors.** *Genome Biol* 2007, **8**:R76.
14. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, et al: **Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts.** *Genome Biol* 2013, **14**:R125.
 15. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L, Karginova O, Jordan J, Combest A, Bridges A, et al: **Predicting Drug Responsiveness in Human Cancers Using Genetically Engineered Mice.** *Clin Cancer Res* 2013, **19**:4889-4899.
 16. Roberts PJ, Usary JE, Darr DB, Dillon PM, Pfefferle AD, Whittle MC, Duncan JS, Johnson SM, Combest AJ, Jin J, et al: **Combined PI3K/mTOR and MEK inhibition provides broad antitumor activity in faithful murine cancer models.** *Clin Cancer Res* 2012, **18**:5290-5303.
 17. Bennett CN, Tomlinson CC, Michalowski AM, Chu IM, Luger D, Mittereder LR, Aprelikova O, Shou J, Piwinica-Worms H, Caplen NJ, et al: **Cross-species genomic and functional analyses identify a combination therapy using a CHK1 inhibitor and a ribonucleotide reductase inhibitor to treat triple-negative breast cancer.** *Breast Cancer Res* 2012, **14**:R109.
 18. Roberts PJ, Bisi JE, Strum JC, Combest AJ, Darr DB, Usary JE, Zamboni WC, Wong KK, Perou CM, Sharpless NE: **Multiple roles of cyclin-dependent kinase 4/6 inhibitors in cancer therapy.** *J Natl Cancer Inst* 2012, **104**:476-487.
 19. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, et al: **Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.** *Nat Med* 2009, **15**:907-913.
 20. Shehata M, Teschendorff A, Sharp G, Novcic N, Russell A, Avril S, Prater M, Eirew P, Caldas C, Watson CJ, Stingl J: **Phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland.** *Breast Cancer Res* 2012, **14**:R134.
 21. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, Perou CM: **Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes.** *Breast Cancer Res Treat* 2013, **142**:237-255.
 22. 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.

23. Spike BT, Engle DD, Lin JC, Cheung SK, La J, Wahl GM: **A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer.** *Cell Stem Cell* 2012, **10**:183-197.
24. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, Marron JS: **Adjustment of systematic microarray data biases.** *Bioinformatics* 2004, **20**:105-114.
25. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *PNAS* 2001, **98**:5116-5121.
26. Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, Anders C, Ewend M, Perou CM: **Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse.** *Breast Cancer Res Treat* 2012, **132**:523-535.
27. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**:346-352.
28. Rousseeuw PJ: **Silhouettes - a Graphical Aid to the Interpretation and Validation of Cluster-Analysis.** *J Comput Appl Math* 1987, **20**:53-65.
29. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, et al: **Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.** *Cancer Cell* 2010, **17**:98-110.
30. 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.
31. Smith GH, Gallahan D, Diella F, Jhappan C, Merlino G, Callahan R: **Constitutive expression of a truncated INT3 gene in mouse mammary epithelium impairs differentiation and functional development.** *Cell Growth Differ* 1995, **6**:563-577.
32. Chan SR, Vermi W, Luo J, Lucini L, Rickert C, Fowler AM, Lonardi S, Arthur C, Young LJ, Levy DE, et al: **STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas.** *Breast Cancer Res* 2012, **14**:R16.
33. Adams JR, Xu K, Liu JC, Agamez NM, Loch AJ, Wong RG, Wang W, Wright KL, Lane TF, Zacksenhaus E, Egan SE: **Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation.** *Cancer Res* 2011, **71**:2706-2717.

34. Li Z, Tognon CE, Godinho FJ, Yasaitis L, Hock H, Herschkowitz JI, Lannon CL, Cho E, Kim SJ, Bronson RT, et al: **ETV6-NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex.** *Cancer Cell* 2007, **12**:542-558.
35. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, et al: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**:1160-1167.
36. Meacham CE, Morrison SJ: **Tumour heterogeneity and cancer cell plasticity.** *Nature* 2013, **501**:328-337.
37. Prater MD, Petit V, Russell IA, Giraddi RR, Shehata M, Menon S, Schulte R, Kalajzic I, Rath N, Olson MF, et al: **Mammary stem cells have myoepithelial cell properties.** *Nat Cell Biol* 2014, **16**:942-950.
38. Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, Cole MD: **A Core MYC Gene Expression Signature Is Prominent in Basal-Like Breast Cancer but Only Partially Overlaps the Core Serum Response.** *PLOS One* 2009, **4**.
39. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, et al: **Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes.** *PNAS* 2010, **107**:15449-15454.
40. Morel AP, Hinkal GW, Thomas C, Fauvet F, Courtois-Cox S, Wierinckx A, Devouassoux-Shisheboran M, Treilleux I, Tissier A, Gras B, et al: **EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice.** *PLOS Genet* 2012, **8**:e1002723.
41. Gluck S, Ross JS, Royce M, McKenna EF, Jr., Perou CM, Avisar E, Wu L: **TP53 genomics predict higher clinical and pathologic tumor response in operable early-stage breast cancer treated with docetaxel-capecitabine +/- trastuzumab.** *Breast Cancer Res Treat* 2012, **132**:781-791.
42. Herschkowitz JI, He X, Fan C, Perou CM: **The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas.** *Breast Cancer Res* 2008, **10**:R75.
43. Prat A, Lluch A, Albanell J, Barry WT, Fan C, Chacon JI, Parker JS, Calvo L, Plazaola A, Arcusa A, et al: **Predicting response and survival in chemotherapy-treated triple-negative breast cancer.** *Br J Cancer* 2014, **111**:1532-1541.

44. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, Smith AE, Prat A, Perou CM, Gilmore H, et al: **Defining the cellular precursors to human breast cancer.** *PNAS* 2012, **109**:2772-2777.
45. Hoadley KA, Weigman VJ, Fan C, Sawyer LR, He X, Troester MA, Sartor CI, Rieger-House T, Bernard PS, Carey LA, Perou CM: **EGFR associated expression profiles vary with breast tumor subtype.** *BMC Genomics* 2007, **8**:258.
46. Efron B, Tibshirani R: **On testing the significance of sets of genes.** *Annals of Applied Statistics* 2007, **1**:107-129.
47. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, Vidaurre T, Holmes F, Souchon E, Wang H, et al: **A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer.** *JAMA* 2011, **305**:1873-1881.
48. Miyake T, Nakayama T, Naoi Y, Yamamoto N, Otani Y, Kim SJ, Shimazu K, Shimomura A, Maruyama N, Tamaki Y, Noguchi S: **GSTP1 expression predicts poor pathological complete response to neoadjuvant chemotherapy in ER-negative breast cancer.** *Cancer Sci* 2012, **103**:913-920.
49. Horak CE, Pusztai L, Xing G, Trifan OC, Saura C, Tseng LM, Chan S, Welcher R, Liu D: **Biomarker analysis of neoadjuvant doxorubicin/cyclophosphamide followed by ixabepilone or Paclitaxel in early-stage breast cancer.** *Clin Cancer Res* 2013, **19**:1587-1595.

CHAPTER 4: SECONDARY GENETIC ABERRATION PROFILING OF P53NULL MAMMARY TUMORS HIGHLIGHTS *MET* DNA AMPLIFICATION AS A GENETIC DRIVER OF MURINE BASAL-LIKE TUMORIGENESIS

OVERVIEW

Background

Breast cancer is the second leading cause of cancer related deaths in American women. Patient care is complicated by inherent tumor heterogeneity that can be classified into at least six intrinsic subtypes. While targeted treatments are standard of care for most subtypes, there remains a clinical need for targeted therapies against basal-like tumors that are typically ‘triple negative breast cancers’. As such, the molecular mechanisms underlying basal-like tumors are under intense investigation to identify genetic drivers and possible drug targets of this subtype.

Methods

Somatic p53 mutations are one of the most common genetic events in basal-like breast tumors. This genetic foundation primes cells for secondary genetic aberration accumulation, a subset of which are predicted to promote tumorigenesis. To identify additional potential drivers of basal-like tumors, a comparative study between human and murine tumors was performed utilizing a p53null mammary transplant murine model. Microarray and sequencing technologies were used to interrogate the secondary genetic aberrations of these murine tumors to then be compared to human basal-like tumors to highlight conserved features.

Results

The p53null mammary murine model produced a genomically diverse set of tumors, a subset of which we showed resemble the human basal-like subtype. Of the ‘omic datasets analyzed, DNA copy number variation produced the largest number of conserved candidate driver genes. Similar to human basal-like tumors, DNA amplification of *Met* was common to murine p53null basal-like tumors. Inhibition of Met using Crizotinib caused these tumors to regress, confirming that this genetic event is a driver of murine basal-like tumorigenesis.

Conclusions

This study identifies Met as a driver of basal-like murine tumors, and thus as a potential driver of human basal-like breast cancer. This work also highlights the importance of comparative genomic studies for discovering novel drug targets and for determining which patient populations are most likely to respond to selective targeted treatments.

BACKGROUND

Breast cancer is a heterogeneous disease that can be segregated into at least six distinct intrinsic subtypes based on gene expression profiles: basal-like, claudin-low, HER2-enriched, luminal A, luminal B, and normal-like [1-3]. While targeted therapeutics exist for estrogen receptor positive [4] (luminal A/B [5]) and human epidermal growth factor receptor 2 positive [6] (HER2-enriched [5]) tumors, targeted treatments for triple negative breast cancer (TNBC) (basal-like and claudin-low [5]) remain an important unmet clinical need [7]. To address this, a research emphasis has been placed on identifying the molecular drivers of basal-like and claudin-low tumors to therefore be exploited as novel drug targets for these subtypes.

Somatic p53 mutations are one of the most frequent genetic events in breast cancer, occurring in about 80% of TNBC [3]. While there is a growing appreciation for the consequences that p53 gain-of-function mutations impose on cell signaling [8], the majority of these mutations are predicted to lead to p53 loss-of-function [9]. This genetic foundation primes tumors for secondary genetic aberration accumulation by decreasing the cell's ability to maintain normal cell physiology. Identifying the subset of genetic events that promote breast cancer is important for informing tumor biology and for guiding personalized treatment regimens. However, segregating genetic drivers of tumorigenesis from passengers is inherently difficult due to the diversity of breast tumors and the large number of candidate aberrations identified in genome-wide profiling studies [3, 10].

Comparative studies between human and murine tumors provide an attractive approach for narrowing the genetic driver candidate list by highlighting conserved features between species [11]. The p53null mammary transplant model [12] is a particularly powerful resource for identifying the genetic drivers of TNBC. From a genetics perspective, the p53null transplant model mimics the loss-of-function seen in human tumors through the expression of a truncated version of p53 [13]. In addition, tumors from this model resemble multiple human intrinsic subtypes of breast cancer, including both basal-like and claudin-low [11, 14]. Identifying the genetic mechanisms that explain this intramodel tumor heterogeneity may help inform the etiology of specific human subtypes. From an experimental perspective, the transplantability of these tumors allows for a single tumor to be expanded and exhaustively studied to verify that the conserved candidates are drivers of tumorigenesis and/or to rigorously test therapeutics [15-18]. For these reasons, this study used the p53null mammary transplant model to identify genetic drivers, and thus novel drug targets, of basal-like breast tumors.

RESULTS

p53null transplant tumors are counterparts for human basal-like and claudin-low subtypes

The p53null transplant model produces phenotypically and genomically diverse tumors that can be classified into three major subtypes/classes based on gene expression profiles: p53null-Basal^{Ex}, Claudin-low^{Ex}, and p53null-Luminal^{Ex} [11]. A critical component of breast cancer comparative studies is to properly identify corresponding human-to-murine subtype counterparts. Once counterparts are determined, conserved features can be identified to highlight the candidate genetic drivers that are specific to those subtypes. For this purpose, we used several approaches to determine which human subtype each of these three p53null transplant genomic classes best represent.

First, a transcriptomic comparison between the two species was performed. To do this, we created gene signatures for each of our three previously identified p53null transplant classes using a two-class (class X versus all others) Significance Analysis of Microarrays (SAM) analysis across a 385 sample microarray dataset with 27 murine models of mammary carcinoma and normal mammary tissue [11]. Each signature was defined as all genes highly expressed in the class of interest with a false discovery rate (FDR) of 0%. The average of these signatures was calculated within each sample of the UNC308 [2], Combined855 [19], and Metabric [10] human breast cancer datasets to identify which human tumors also highly expressed these same set of genes. Interestingly, both the p53null-Basal^{Ex} and p53null-Luminal^{Ex} signatures were highly expressed in basal-like human tumors, while the Claudin-low^{Ex} signature was most specific to the human claudin-low subtype (Figure 13A). Similar results were observed when we compared molecular pathway based signatures (Figure 14A) [11].

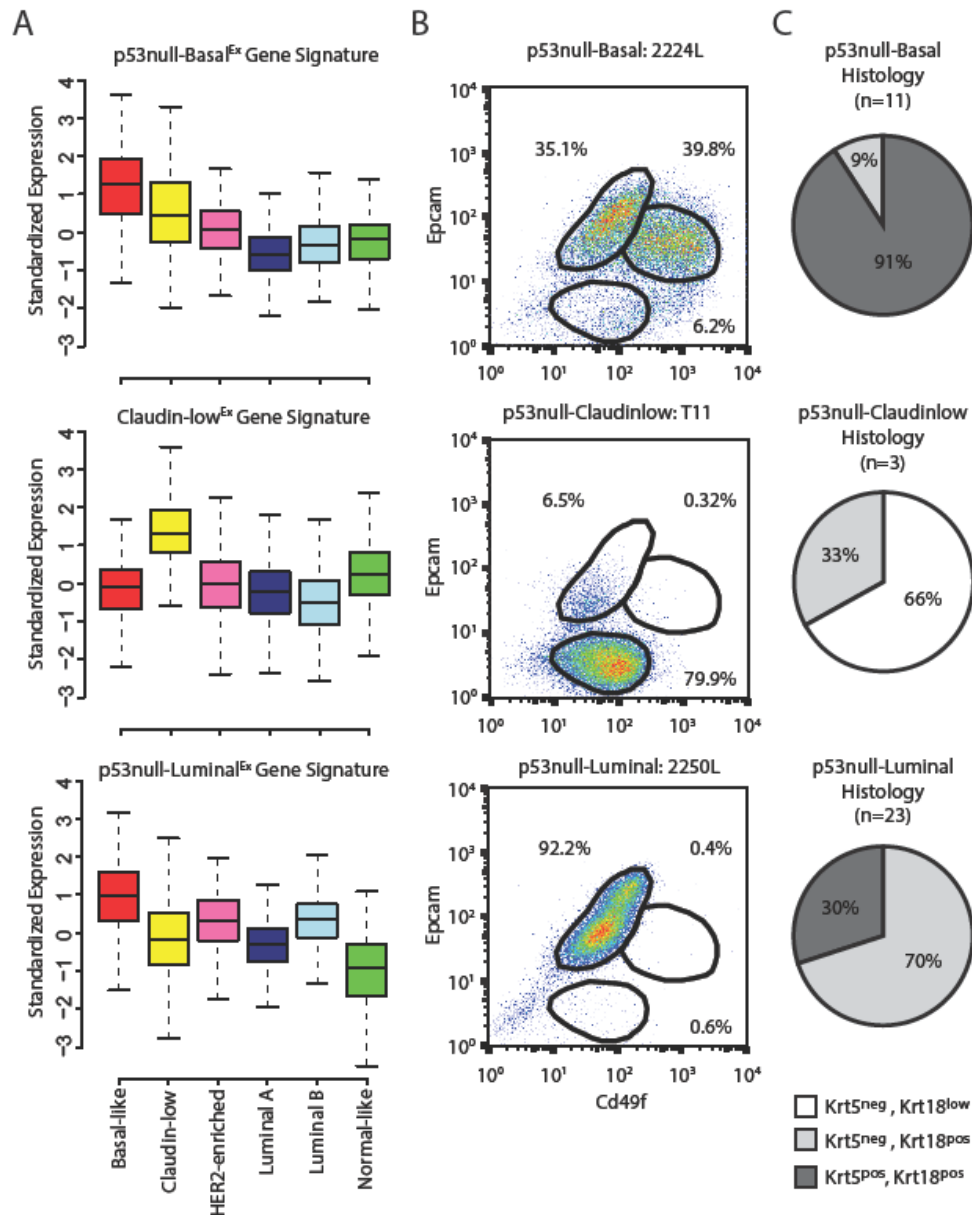


Figure 13: Human counterparts of p53null transplant tumors

A. Genes highly expressed within each p53null transplant class were identified using a 2-class (class X vs all others) SAM analysis (FDR 0%) across our 385 murine microarray dataset. The standardized average of these gene signatures was calculated across more than three thousand human tumors and displayed by intrinsic subtype. **B.** A representative tumor from each p53null transplant class was FACS sorted using antibodies against Epcam and Cd49f. **C.** Immunohistochemistry was performed using antibodies against Krt5 (a marker of basal cells) and Krt18 (a marker of luminal cells).

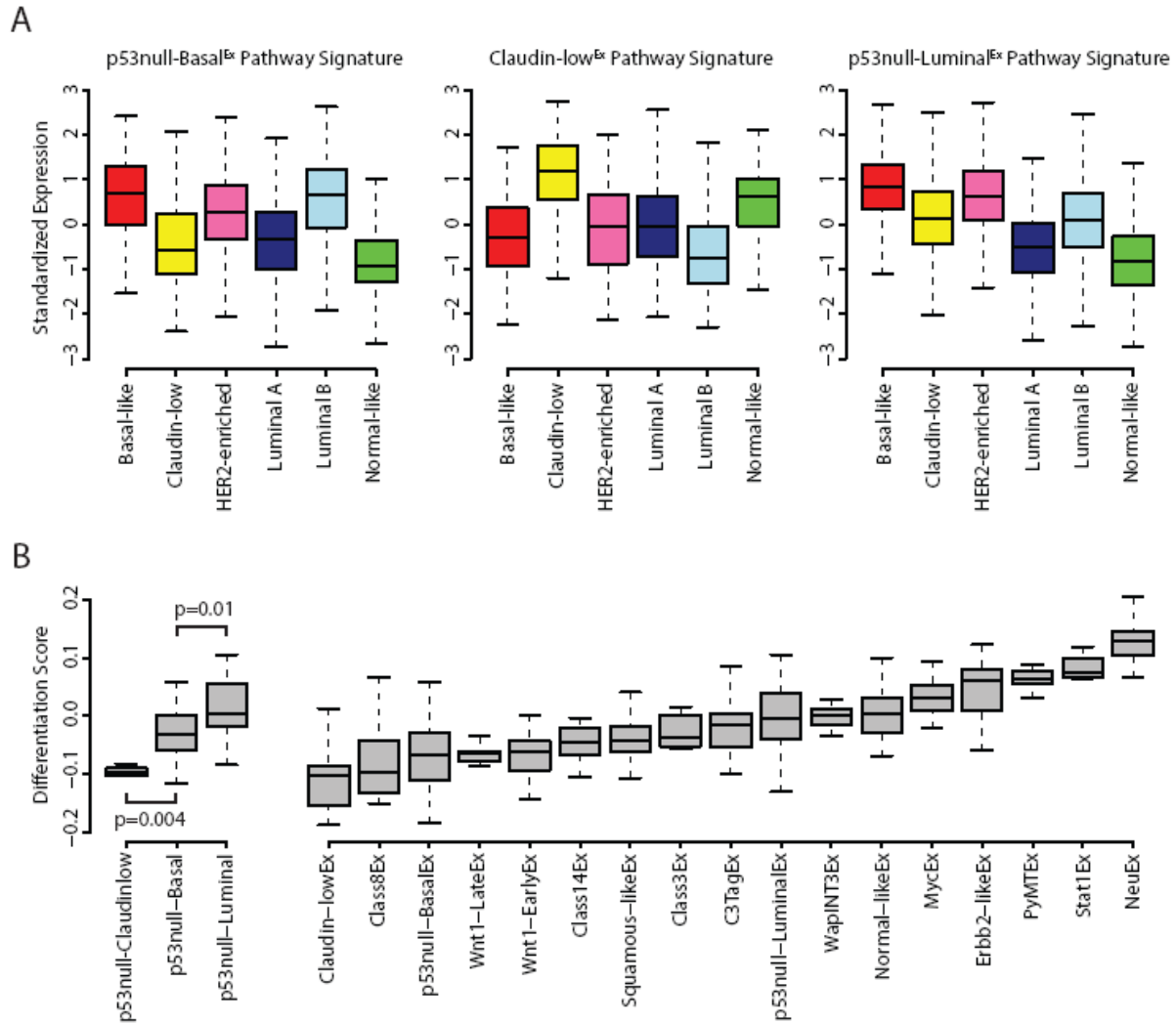


Figure 14: Molecular pathway signatures and differentiation score

A. Median expression values were calculated for 963 publicly available pathway-based gene signatures across our murine and human expression datasets [11]. Molecular pathway signatures were identified using a 2-class (class X vs all others) SAM analysis (FDR 0%) across our 385 murine microarray dataset. The standardized average of these gene signatures was calculated across more than three thousand human tumors and displayed by intrinsic subtype. **B.** Tumor ‘differentiation scores’ [2] were calculated for all 385 murine samples and displayed by intrinsic class. The ‘differentiation scores’ of the three p53null transplant classes were compared using a t-test.

One explanation for the transcriptomic associations observed in Figure 13A is that both the human and murine subtypes arise from similar cell types within the mammary gland [20]. To address this possibility, a representative tumor from each p53null transplant subtype was selected and fluorescence-activated cell sorted (FACS) into its mammary subpopulations using Cd49f and Epcam markers [21-23] (Figure 13B). The 2224L (p53null-Basal^{Ex}) tumor contained a mixture of epithelial cell subpopulations, with 35% of cells being Cd49f^{pos}/Epcam^{pos} and 40% of cells being Cd49f^{double pos}/Epcam^{pos} (Figure 13B). The T11 (p53null-Claudinlow^{Ex}) and 2250L (p53null-Luminal^{Ex}) tumors, however, dissociated into single epithelial subpopulations. Specifically, 80% of T11 (p53null-Claudinlow^{Ex}) epithelial cells were distinctly Cd49f^{pos}/Epcam^{neg}. Expression profiles of normal human mammary CD49f^{pos}/EpCAM^{neg} subpopulations are enriched within the human claudin-low subtype, indicating that human and mouse claudin-low tumors share similar FACS-based cellular features [24]. Greater than 90% of 2250L (p53null-Luminal^{Ex}) epithelial cells were Cd49f^{pos}/Epcam^{pos}. Human basal-like tumors similarly share features of CD49f^{pos}/EpCAM^{pos} epithelial cells [23, 24], but tend to show a more mixed cellular population as was seen in the murine 2224L (p53null-Basal^{Ex}) tumor.

While these FACS profiles are informative, only a subset of our p53null transplant tumors have been properly processed for passaging experiments. As such, we were restricted from broadly expanding this approach to all of the tumors in our dataset. To circumvent this limitation, a ‘differentiation score’ (D-Score) was calculated for all tumors in the murine microarray dataset (Figure 14B) [2]. Low scores indicate a tumor similarity to adult mammary stem cells (aMaSC), intermediate scores a similarity to luminal progenitor (LumProg) cells, and high scores a similarity to mature luminal (MatureLum) cells [2]. Similar to the individual tumor FACS profiles, the D-Score varied across the three p53null subtypes, with the p53null-

Claudin^{low}^{Ex} subtype being the lowest, the p53null-Basal^{Ex} being intermediate, and the p53null-Luminal^{Ex} being the highest ($p < 0.05$) (Figure 14B). Even though the p53null-Luminal^{Ex} subtype had the highest D-Score among the three p53null subtypes, its score is still relatively intermediate when compared across our diverse murine tumor dataset. This indicates that while the p53null-Luminal^{Ex} class is the most ‘luminal’ of the three p53null classes, these tumors do not have as strong an association to MatureLum cells as do human luminal A and luminal B tumors [24] and murine MMTV-Neu tumors.

To supplement these findings, a histological characterization of the cellular features of p53null transplant mammary tumors was also performed. Specifically, we stained tumors with Krt5 (a marker of basal cells) and Krt18 (a marker of luminal cells). Consistent with the FACS profile (Figure 13B), 91% of p53null-Basal^{Ex} tumors were Krt5^{pos}/Krt18^{pos}, indicating that these tumors contain features of both basal and luminal cell types. On the contrary, 66% of p53null-Claudin^{low}^{Ex} tumors were Krt5^{neg}/Krt18^{low}. Consistent with their nomenclature, p53null-Luminal^{Ex} tumors primarily contained luminal cell type features, with 70% of tumors being Krt5^{neg}/Krt18^{pos}. Taken together (Figure 13), these results indicate that the p53null transplant model produces tumors that are best considered murine counterparts for human basal-like and claudin-low tumors.

Secondary genetic aberration profiling highlights DNA copy number changes as drivers of tumorigenesis

In broad terms, disruption of normal p53 signaling leads to an unstable genome due to a decreased ability to properly respond to the presence of genetic aberrations [25]. This phenotype leads to the accumulation of both small scale mutations (e.g. insertions, deletions) and large scale

chromosomal rearrangements (e.g. translocations, copy number variations) throughout the genome. Specific genetic aberrations are predicted to be responsible for determining a cell's subtype fate during tumorigenesis, but identifying these specific drivers has been challenging. Here, we leveraged the power of multiple 'omic technologies to interrogate the secondary genetic aberrations underlying 43 different and independently arisen p53null transplant tumors. Specifically, microarray and sequencing technologies were used to produce five datasets of varying sizes: DNA exome analysis (EXO) (n=37), DNA copy number (n=43), chromosome structural variation (n=37), expression microarrays (n=43), and RNAseq (n=6) (Figure 15).

Given our hypothesis that the accumulation of specific secondary genetic events during tumorigenesis drives the development of specific tumor subtypes, we designed our statistically analyses to identify those genetic events that are enriched within specific p53null transplant classes as compared to the other two. Using this approach with our DNA EXO dataset, there were no identifiable genes that were somatically mutated and enriched within any of the three p53null transplant classes using a 2-class (class X versus all others) fisher's exact test (Figure 16). While this was initially a somewhat surprising result, it is consistent with The Cancer Genome Atlas's (TCGA) profiling of human breast tumors in which there was only one gene with a mutation frequency greater than ten percent within the basal-like subtype (*TP53*) [3].

Given the mutation results, we decided to focus on large scale chromosomal rearrangements, amplifications, and/or deletions as possible drivers of the p53null transplant subtypes. An analysis of chromosome structural variants identified several rearrangements that were enriched within each of the three p53null transplant classes using a 2-class (class X versus all others) fisher's exact test (Figure 17). For instance, *Mad2l1* was enriched for structural rearrangements in p53null-Basal^{Ex} tumors. *Mad2l1* plays an important role during metaphase by

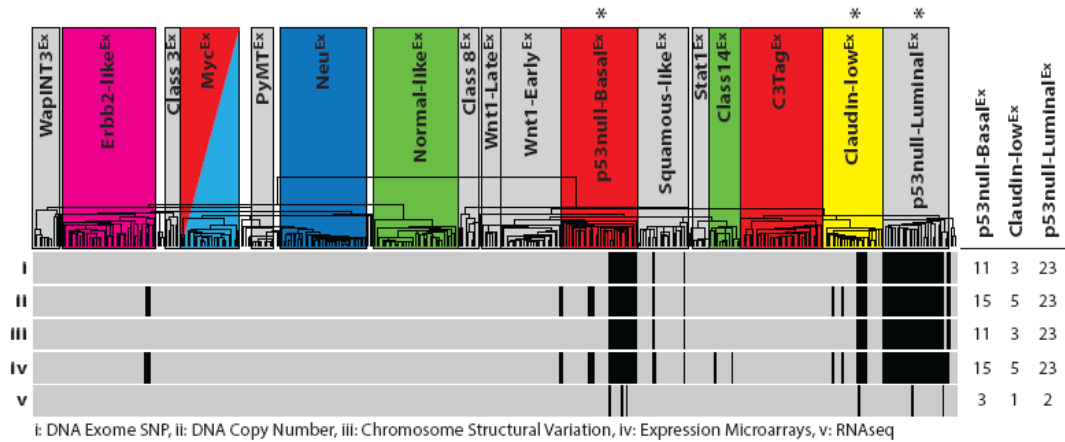


Figure 15: Murine p53null tumor datasets

Microarray and sequencing technologies were used to produce five p53null tumor datasets of varying sizes: **i.** DNA SNP (n=37), **ii.** DNA copy number (n=43), **iii.** Chromosome structural variation (n=37), **iv.** Expression microarrays (n=43), **v.** and RNAseq (n=6). The intrinsic class of each sample is displayed on the dendrogram, with colored boxes being previously identified human subtype counterparts [11]. The hierarchical clustering location of each p53null tumor within the datasets is displayed as a vertical black strip.

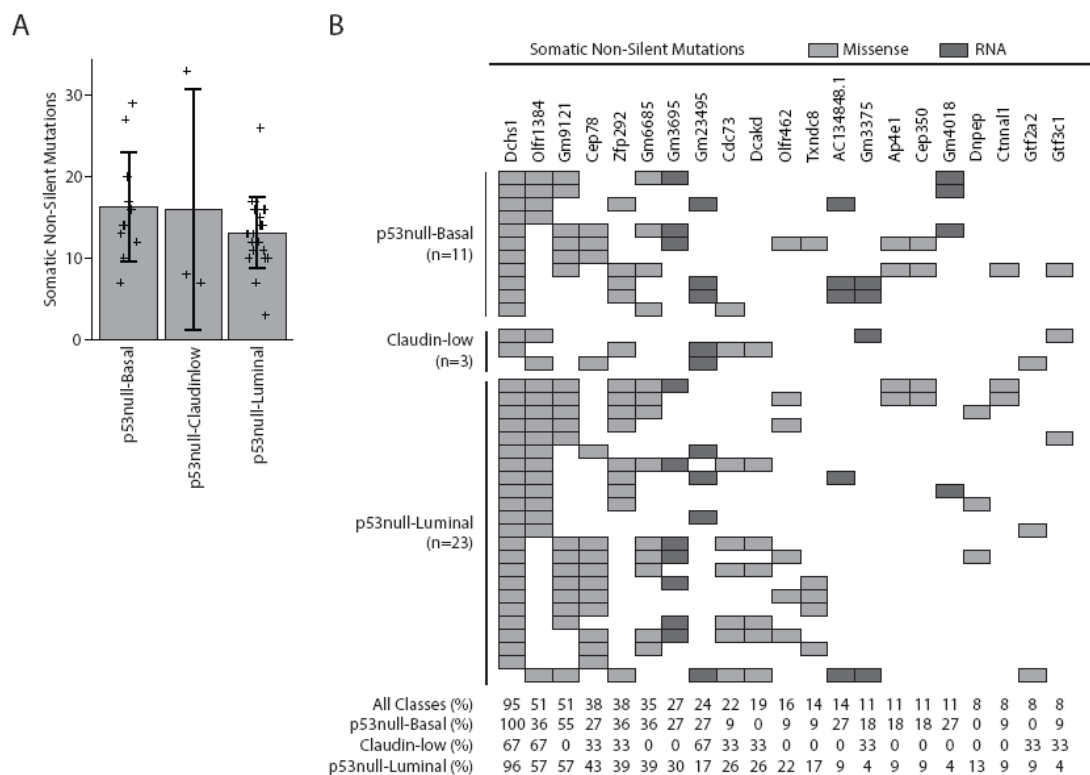


Figure 16: Somatic non-silent mutation analysis

A. The total number of somatic non-silent mutations was determined for each sample and plotted by p53null transplant class. Error bars represent one standard deviation. **B.** All somatic non-silent mutations with a mutational frequency greater than or equal to 8% across all p53null tumors are displayed by p53null class. Each row represents an individual tumor, with boxes corresponding to a mutation within the gene on that column.

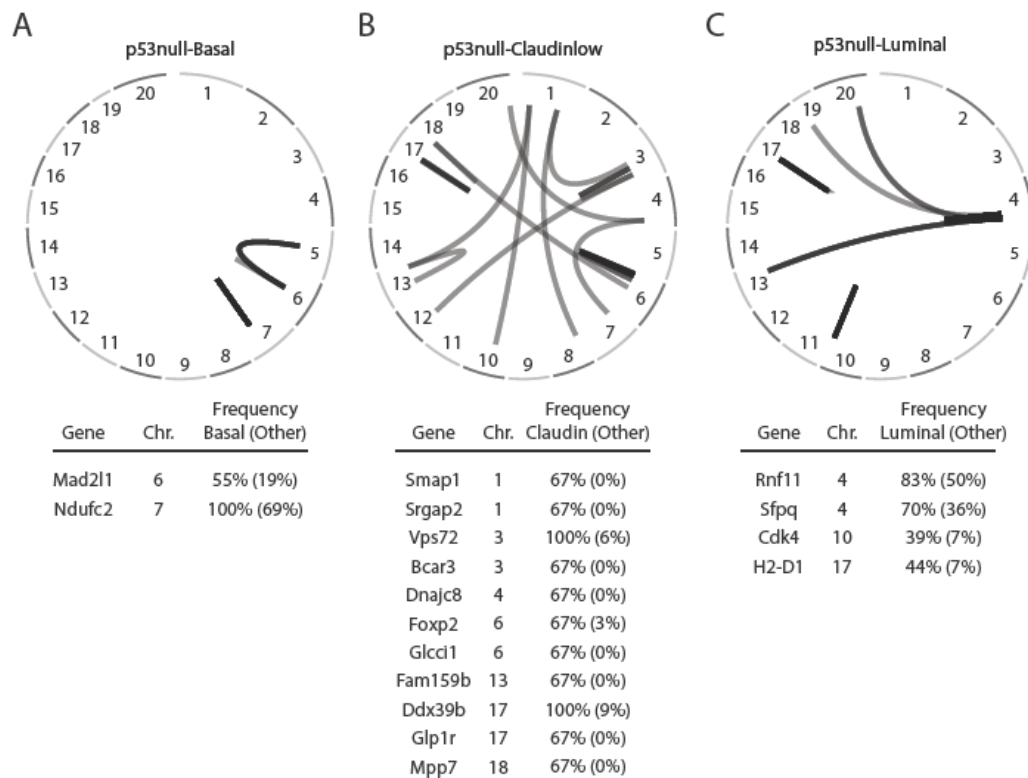


Figure 17: Chromosome structural variation analysis

Displayed are circos plots of the structural variants enriched within **A.** p53null-Basal, **B.** p53null-Claudinlow, and **C.** p53null-Luminal tumors as determined by 2-class (class X versus all others) Fisher's Exact Tests.

preventing progression into anaphase until all chromosomes are properly aligned [26].

Overexpression of *Mad21l*, as is the case in p53null-Basal^{Ex} tumors (FDR 0%), may further promote tumor development by decreasing chromosome stability [27]. While these p53null transplant class enriched structural variants are intriguing, it is inherently difficult to perform follow-up experiments on these genetic events to gauge their effect on the tumor phenotype. Thus, we are unable to definitively call any of these structural variants drivers of these classes.

The mechanism by which DNA copy number variation leads to changes in the tumor phenotype, however, is more intuitive and more easily tested; therefore, we decided to focus our attention on these amplifications and/or deletions as secondary genetic drivers. We were primarily interested in identifying genes in which their DNA copy number variation was highly correlated with their gene expression, as this observation is consistent with causality. First, we identified DNA copy number changes enriched within each of our three p53null transplant classes using a 2-class (class X versus all others) SAM analysis (Figure 18). Interesting, both the p53null-Basal^{Ex} (Figure 18A) and p53null-Luminal^{Ex} (Figure 18C) classes had distinct genomic regions of DNA gains and losses, while the p53null-Claudinlow^{Ex} (Figure 18B) class was more copy number neutral, having no genomic regions enriched with gains or losses. These results are consistent with human studies which have highlighted several DNA copy number events specific to basal-like tumors but few, if any, events in claudin-low tumors [28]. Specifically, p53null-Basal^{Ex} tumors were defined by both gains and losses on chromosome 8 and almost a complete loss of chromosome 12 (Figure 18A), while p53null-Luminal^{Ex} tumors were defined by a DNA amplification on chromosome 6 (Figure 18C).

To narrow the list of potential genetic drivers within these subtype-specific regions of gains and losses, the Pearson correlation between DNA copy number and gene expression was

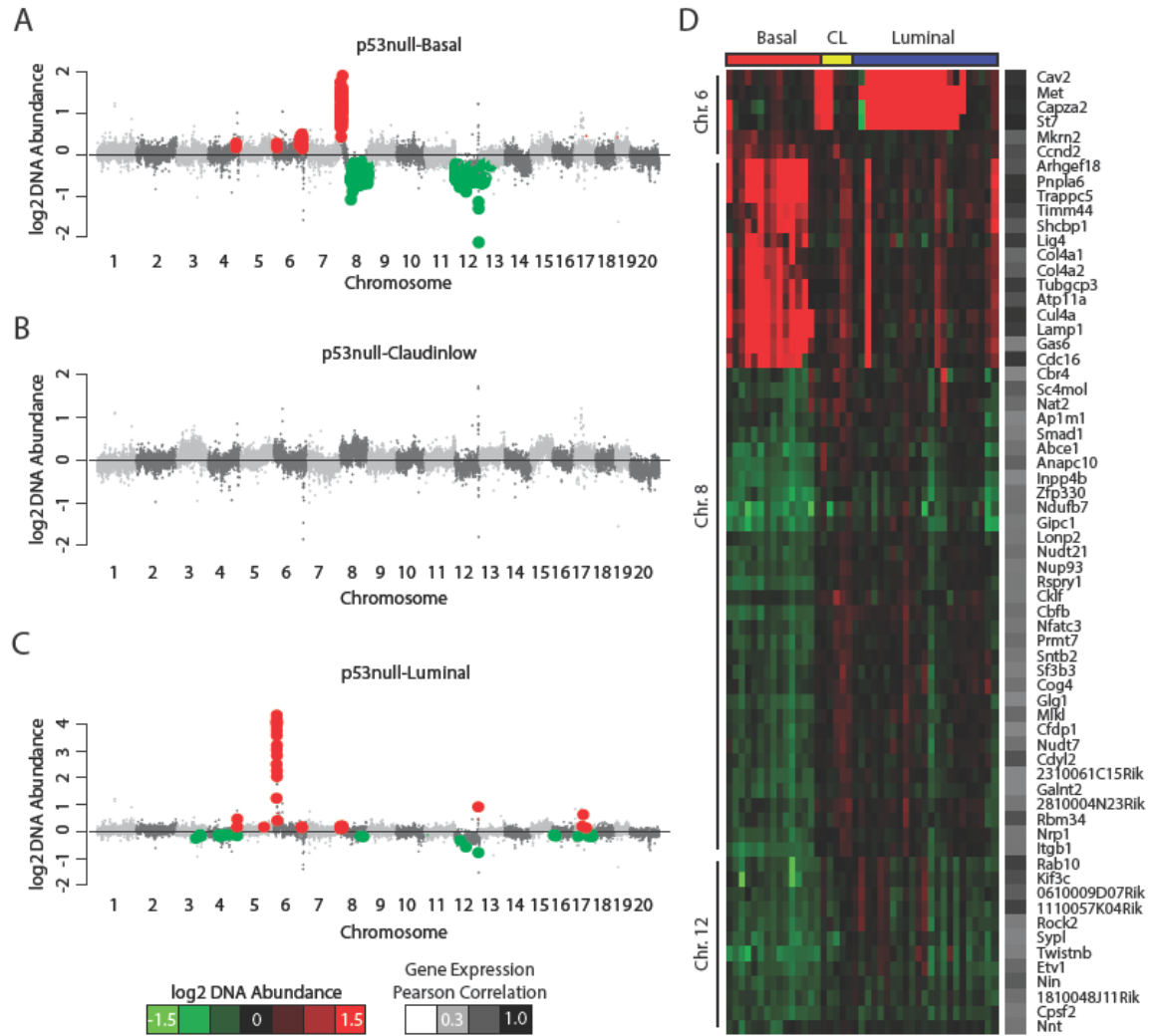


Figure 18: DNA copy number analysis

Displayed in genomic order are the median class DNA copy number levels for **A.** p53null-Basal, **B.** p53null-Claudinlow, and **C.** p53null-Luminal tumors. DNA copy number changes enriched within each of the three p53null transplant classes were identified using a 2-class (class X versus all others) SAM analysis. Genomic regions of significant gain are labeled in red and regions of significant loss are labeled in green. **D.** Pearson correlations between DNA copy number and gene expression were determined for all genes within the significant regions of gain and loss from parts A-C. Genes with a correlation greater than or equal to 0.5 are displayed in genomic order. The heatmap corresponds to DNA copy number abundance.

calculated to highlight the genes that are the most sensitive to DNA gains and losses. In this case, genes with Pearson correlations greater than 0.5 were considered to have a strong association (Figure 18D). A number of interesting genes fell into this classification. For instance, *Inpp4b*, a regulator of PI3K/AKT signaling [29], was lost in p53null-Basal^{Ex} tumors on chromosome 12, similar to human basal-like tumors [28]. *Cul4a*, which is located on the p53null-Basal^{Ex} chromosome 8 amplicon, had a very high correlation with its gene expression (Pearson=0.86) (Figure 18D and Figure 19A). CUL4A is a scaffolding protein for E3 ubiquitin ligase that helps to regulate the cellular concentration of key protein substrates, including CHK1, E2F1, ER- α , and pol η to name a few [30]. Given the wide variety of cellular phenotypes that these protein substrates influence, such as proliferation and DNA repair [30], CUL4A has been proposed to be an attractive cancer drug target [31]. CUL4A amplification and overexpression has been observed in the human basal-like breast cancer subtype and has been demonstrated to be a driver of tumorigenesis both in-vitro and in-vivo [32-34]. p53null-Luminal^{Ex} tumors had a distinct amplification of chromosome 6 (Figure 18C). Within this region, six genes had high correlation with their gene expression (Figure 18D), including *Met* (Pearson=0.92) (Figure 19B). MET is a receptor tyrosine kinase for hepatocyte growth factor that regulates a variety of downstream signal transduction pathways, including MAPK and PI3K/AKT [35].

Given that both p53null-Basal^{Ex} and p53null-Luminal^{Ex} tumors are counterparts for human basal-like tumors (Figure 13), we investigated the TCGA DNA copy number dataset to see if these genetic events also frequently occur in human breast tumors. Both MET and CUL4A were amplified in ~20% of human basal-like tumors (Table 10). Interestingly, these genetic events generally did not co-occur within the same tumor, similar to our p53null mouse model. Given that Met DNA amplification is a conserved feature of human and mouse basal-like

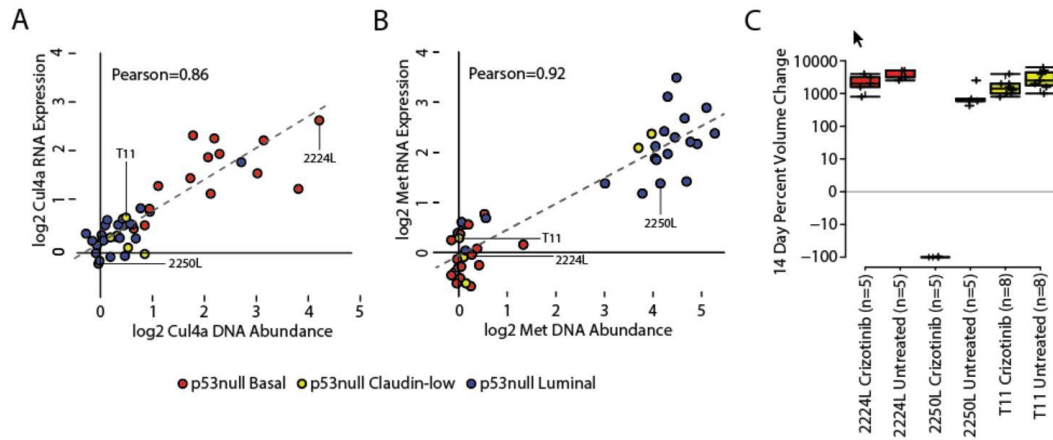


Figure 19: *MET* DNA amplification is a driver of 2250L (p53null-Luminal^{Ex}) tumors
A. Pearson correlation between *Cul4a* DNA copy number and gene expression **B.** Pearson correlation between *Met* DNA copy number and gene expression **C.** p53null mammary transplant tumor response to Crizotinib treatment

Human Subtype	MET Amplified	CUL4A Amplified	Co-Amplified
Basal-like	20.3%	22.0%	6.8%
HER2-enriched	3.3%	8.3%	0.0%
Luminal A	4.8%	3.4%	0.6%
Luminal B	4.7%	6.0%	0.6%
Normal-like	0.0%	6.7%	0.0%
Total	7.2%	7.6%	1.5%

Table 10: DNA amplification status of *CUL4A* and *MET* across the TCGA breast cancer dataset

2250L (p53null-Luminal^{Ex}) tumors completely regress with Crizotinib treatment

Similar to *Cul4a*, *Met* DNA copy number had a very high correlation with its gene expression (Figure 19B). To determine if *Met* is a driver in *Met* amplified tumors, noting that this gene was not somatically mutated, we treated three p53null transplant tumor lines with Crizotinib, a US Food and Drug Administration (FDA) approved therapy for non-small-cell lung cancer [36] that inhibits *Met* and ALK [37]. While neither 2224L (p53null-Basal^{Ex}) nor T11 (p53null-Claudinlow^{Ex}) tumors responded to treatment (noting that both of these lines were not *MET* amplified), all of the 2250L (p53null-Luminal^{Ex}) tumors had complete regression at the end of the 14 day treatment period (Figure 19C). Since *Alk* is not differentially expressed across the three p53null transplant classes, we propose that this dynamic response is due to differences in *Met* signaling, and thus, *Met* amplification is a driver of tumorigenesis in this mouse model of basal-like breast cancer.

DISCUSSION

Even though increased public awareness and a greater understanding of tumor biology have led to improved patient survival rates, breast cancer is still the second leading cause of cancer related deaths in American women. With so many patients either not responding or relapsing with the current standard of care, the molecular mechanisms underlying breast cancer are under intense investigation to identify new, personalized drug targets [38, 39]. This is especially true for triple negative breast cancer (TNBC) (basal-like and claudin-low) for which targeted treatment options remain an important unmet clinical need. Murine models provide an excellent resource for identifying genetic drug targets by highlighting conserved features between species [11]. Given that somatic p53 mutations are one of the most common genetic

events in TNBC [3], the p53null mammary transplant mouse model is particularly useful for studying the molecular mechanisms of TNBCs.

The p53null transplant model produces heterogeneous tumors that can be classified into three major subtypes/classes based on gene expression profiles: p53null-Basal^{Ex}, Claudin-low^{Ex}, and p53null-Luminal^{Ex} [11]. Using a combination of gene expression comparisons, FACS analysis, and immunohistochemistry, we show that p53null-Basal^{Ex} and p53null-Luminal^{Ex} tumors are counterparts for the human basal-like subtype, while p53null-claudinlow^{Ex} tumors are counterparts for the human claudin-low subtype. Even though p53null-Luminal^{Ex} tumors were the most ‘luminal’ of the p53null classes, these tumors were found to more closely resemble luminal progenitor cells than mature luminal cells. Specifically, FACS analysis of the 2250L (p53null-Luminal^{Ex}) line indicated that this tumor is Cd49f^{pos}/Epcam^{pos}, the same FACS profile as luminal progenitor cells [23]. In addition, p53null-Luminal^{Ex} tumors have intermediate ‘differentiation scores’, similar to luminal progenitor cells [2]. The original nomenclature for this class was derived by an observed association to the luminal subtypes based on a few luminal markers [14], but recent work has shown that mature luminal and luminal progenitor cells share many of the same features [24, 40]. For instance, both FACS populations are EpCAM^{pos} and Krt18^{pos} [23, 40], indicating that broader analyses are required to distinguish between these cell types within tumors. These findings help explain why p53null-Luminal tumors were found to be counterparts for the basal-like subtype and not the luminal A/B subtypes, as basal-like tumors also share features of luminal progenitor cells [23, 24].

Once these human-murine subtype counterparts were defined, secondary genetic aberration profiling was performed to identify conserved events between species to highlight candidate drivers of tumorigenesis. Of the ‘omic datasets analyzed, DNA copy number variation

produced the largest number of candidate genes. We were particularly interested in genes that had high correlation between their copy number and gene expression because for these cases, we propose that the copy number change is the mechanism that directly influences the expression of the genes within those genomic regions.

The murine p53null-Luminal^{Ex} class showed a reproducible amplification of *MET* (without mutation), thus suggesting that this could be a driving event for this murine subtype. Crizotinib treatment resulted in complete tumor regression (not palpable) at the end of the 14 day treatment period in our 2250L (p53null-Luminal^{Ex}) tumor line. *MET* is an important receptor tyrosine kinase that can activate a variety of signal transduction pathways, including MAPK and PI3K/AKT. These results suggest that *MET* is a driving oncogene in this subtype, and that Crizotinib may perform very well as a single agent against breast tumors with aberrant *MET* signaling. These experimental results are particularly relevant given that about 20% of human basal-like tumors have amplification of *MET*. Since not all basal-like patients have these aberrations, companion diagnostic tests will be required to determine which patients should receive treatment which, in this case, could be based upon the presence or absence of *MET* amplification. Proper clinical studies will be required to determine if that is the case.

In summary, we identified a number of class specific copy number events in our p53null mouse model, which mimicked similar events in human basal-like tumors. Using a comparative genomics approach, this study identified *MET* as a driver of one of these mouse classes and highlights the potential link of *MET* to human basal-like breast cancer [35]. In addition to this clinically important finding, it also highlights the importance of comparative genomic studies as a preclinical tool for discovering novel drug targets and for determining which patient populations are most likely to respond to treatment. Currently, only ~5% of oncology drugs that

enter clinical testing are ultimately approved by the US FDA for use [41]. From a financial perspective, this low success rate means that it cost more than \$1,000,000,000 of research and development spending to develop a single oncology drug [42]. Needless to say, improved methods are needed to streamline the drug development process and to lower healthcare costs. As supported by the results of this study, we propose that mouse models should be an integral part of early phase drug development to highlight those drug/drug combinations most likely to succeed in clinical trials.

MATERIALS AND METHODS

Gene expression

Microarray gene expression data from 27 murine models of mammary carcinoma and normal mammary tissue were downloaded from the following gene expression omnibus (GEO) entries: GSE3165, GSE8516, GSE9343, GSE14457, GSE15263, GSE17916, GSE27101, and GSE42640 [11]. The 385 sample dataset was normalized to correct for microarray platform bias as previously described [11].

Tumor differentiation scores were calculated across the microarray dataset as previously described [2]. Gene expression signatures were created for the three murine classes enriched with the p53null transplant model (p53null-Basal^{Ex}, Claudin-low^{Ex}, and p53null-Luminal^{Ex}) by performing a two-class (class X versus all others) Significance Analysis of Microarrays (SAM) analysis on the microarray dataset [43]. Signatures were defined as all genes highly expressed in the class of interest with a false discovery rate (FDR) of 0%. Similarly, pathway signatures were created as previously described [11]. Expression scores for each gene and pathway signature

were determined by calculating the mean expression of the signature within each sample in the UNC308 [2], Combined855 [19], and Metabric [10] human breast cancer datasets.

Flow cytometry

p53null transplant tumors were dissociated into a single cell suspension using the following steps. First, each tumor was manually cut into small pieces with a razor blade in 1X collagenase/hyaluronidase (StemCell #07919) EpiCult media (StemCell #05601) before being placed in a rotator for two hours at 37°C. Following lysis of red blood cells using ammonia chloride (StemCell #07850), the tumors were incubated in 1X trypsin-EDTA (Sigma #T4049) for five minutes at 37°C and then in a 1X Dispase (StemCell #07923) DNase I solution (StemCell #07900) for five min at 37°C to reduce cell clumping. Cells were filtered through a 40µm nylon cell strainer (Fisher Scientific #08-771-1) in HBSS media (StemCell #37150) with 10% FBS (Sigma #F2442) to obtain the final single cell suspension. To remove non-epithelial cells, the single cell suspension was taking through a mouse epithelial cell enrichment kit (StemCell #19758) following the manufacturer's protocol. Cells were labeled with the following antibodies for 30 minutes at 4°C: FITC anti-mouse Epcam (eBioscience #11-5791-82) and APC anti-mouse Cd49f (eBioscience #17-0495-82). Fluorescence-activated cell sorting (FACS) was performed using a Beckman-Coulter CyAn ADP instrument and analyzed using the FlowJo v10 software program.

DNA single nucleotide polymorphisms

Mutation data was collapsed to a gene level so that all non-silent somatic mutations affecting the same gene were treated equally regardless of the actual mutation. A two-class (class

X versus all others) fisher's exact test (FET) was performed to identify genes preferentially mutated within each p53null class (p-value<0.05).

DNA structural variants

Genomic structural variants (SV) were collapsed to a gene level so that all SV affecting the same gene were treated equally regardless of the actual SV. Genes were defined as being affected by the structural variant if the start or end of the SV occurred within the RefSeq gene region. A two-class (class X versus all others) FET was performed to identify genes preferentially affected by SV within each p53null class (p-value<0.05). Because all of the p53null Claudin-low^{Ex} tumors were analyzed using whole genome sequencing, a second two-class FET was performed on these tumors in which only the 13 whole genome profiled tumors were included in the analysis to reduce the likelihood of the p53null Claudin-low^{Ex} enriched SV being an artifact of methodology.

DNA copy number

DNA array comparative genomic hybridization (aCGH) data was downloaded for the p53null transplant tumors classified as p53null-Basal^{Ex}, Claudin-low^{Ex}, or p53null-Luminal^{Ex} by gene expression profiling from GEO entry GSE27101 [14]. In addition, genomic DNA was extracted from five p53null transplant tumors using a DNeasy blood and tissue kit (Qiagen #69504), labeled with a Sure Tag DNA kit (Agilent #5190-4240), and hybridized to 244K CGH microarrays (Agilent #G4415A) as previously described [14].

The 43 sample aCGH dataset was extracted from the UNC Microarray Database as log₂ Cy5/Cy3 ratios, filtering for probes with Lowess normalized intensity values greater than ten in

the control channel and for probes with data on greater than 70% of the microarrays [14]. The probes that passed these filters were then oriented in genomic order and a ten probe average was calculated on consecutive groups of ten probes across each chromosome, resulting in a final dataset of 23,181 features. A two-class (class X versus all others) SAM analysis was performed to identify genomic regions of amplification or deletion unique to each class (FDR of 0%).

Level 3 DNA segmentation data was downloaded from The Cancer Genome Atlas (TCGA) data portal for 715 breast cancer samples. Genomic regions of amplification were defined as having a \log_2 segmentation value greater than 0.3.

Crizotinib treatment

All mouse work was performed under protocols approved by the UNC Institutional Animal Care and Use Committee (IACUC). One million p53null transplant cells were suspended in matrigel and injected subcutaneously into the mammary pad of BALB/c wild-type female mice. Upon tumor formation, mice were randomized to either the Crizotinib (ChemShuttle #877399-52-5) or untreated group. Crizotinib chow was synthesized by OpenSource Diets to a final concentration of 50 mg/kg/day and was given continuously over the 14 day treatment period to monitor tumor growth. Tumor volume was calculated from two-dimensional measurements ($\text{Volume} = [(\text{width})^2 \times \text{length}] / 2$). The percent change in volume at 14 days was used to quantify response.

REFERENCES

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
2. 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.
3. Cancer Genome Atlas N: **Comprehensive molecular portraits of human breast tumours.** *Nature* 2012, **490**:61-70.
4. Jordan VC: **Tamoxifen: a most unlikely pioneering medicine.** *Nat Rev Drug Discov* 2003, **2**:205-213.
5. Prat A, Perou CM: **Deconstructing the molecular portraits of breast cancer.** *Mol Oncol* 2011, **5**:5-23.
6. Hynes NE, Lane HA: **ERBB receptors and cancer: the complexity of targeted inhibitors.** *Nat Rev Cancer* 2005, **5**:341-354.
7. Carey L, Winer E, Viale G, Cameron D, Gianni L: **Triple-negative breast cancer: disease entity or title of convenience?** *Nat Rev Clin Oncol* 2010, **7**:683-692.
8. Brosh R, Rotter V: **When mutants gain new powers: news from the mutant p53 field.** *Nat Rev Cancer* 2009, **9**:701-713.
9. Bullock AN, Fersht AR: **Rescuing the function of mutant p53.** *Nat Rev Cancer* 2001, **1**:68-76.
10. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**:346-352.
11. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, et al: **Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts.** *Genome Biol* 2013, **14**:R125.
12. Jerry DJ, Kittrell FS, Kuperwasser C, Laucirica R, Dickinson ES, Bonilla PJ, Butel JS, Medina D: **A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development.** *Oncogene* 2000, **19**:1052-1058.
13. Jacks T, Remington L, Williams BO, Schmitt EM, Halachmi S, Bronson RT, Weinberg RA: **Tumor spectrum analysis in p53-mutant mice.** *Curr Biol* 1994, **4**:1-7.

14. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, Knezevic J, Greene SB, Darr D, Troester MA, et al: **Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells.** *Proc Natl Acad Sci U S A* 2012, **109**:2778-2783.
15. Roberts PJ, Bisi JE, Strum JC, Combest AJ, Darr DB, Usary JE, Zamboni WC, Wong KK, Perou CM, Sharpless NE: **Multiple roles of cyclin-dependent kinase 4/6 inhibitors in cancer therapy.** *J Natl Cancer Inst* 2012, **104**:476-487.
16. Roberts PJ, Usary JE, Darr DB, Dillon PM, Pfefferle AD, Whittle MC, Duncan JS, Johnson SM, Combest AJ, Jin J, et al: **Combined PI3K/mTOR and MEK inhibition provides broad antitumor activity in faithful murine cancer models.** *Clin Cancer Res* 2012, **18**:5290-5303.
17. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L, Karginova O, Jordan J, Combest A, Bridges A, et al: **Predicting drug responsiveness in human cancers using genetically engineered mice.** *Clin Cancer Res* 2013, **19**:4889-4899.
18. Bennett CN, Tomlinson CC, Michalowski AM, Chu IM, Luger D, Mittereder LR, Aprelikova O, Shou J, Piwinica-Worms H, Caplen NJ, et al: **Cross-species genomic and functional analyses identify a combination therapy using a CHK1 inhibitor and a ribonucleotide reductase inhibitor to treat triple-negative breast cancer.** *Breast Cancer Res* 2012, **14**:R109.
19. Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, Anders C, Ewend M, Perou CM: **Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse.** *Breast Cancer Res Treat* 2012, **132**:523-535.
20. Visvader JE: **Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis.** *Genes Dev* 2009, **23**:2563-2577.
21. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, Perou CM: **Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes.** *Breast Cancer Res Treat* 2013, **142**:237-255.
22. Shehata M, Teschendorff A, Sharp G, Novcic N, Russell IA, Avril S, Prater M, Eirew P, Caldas C, Watson CJ, Stingl J: **Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland.** *Breast Cancer Res* 2012, **14**:R134.
23. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, et al: **Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.** *Nat Med* 2009, **15**:907-913.

24. Pfefferle AD, Spike BT, Wahl GM, Perou CM: **Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant chemotherapy.** *Breast Cancer Res Treat* 2015, **149**:425-437.
25. Bieging KT, Mello SS, Attardi LD: **Unravelling mechanisms of p53-mediated tumour suppression.** *Nat Rev Cancer* 2014, **14**:359-370.
26. Schvartzman JM, Sotillo R, Benezra R: **Mitotic chromosomal instability and cancer: mouse modelling of the human disease.** *Nat Rev Cancer* 2010, **10**:102-115.
27. Sotillo R, Hernando E, Diaz-Rodriguez E, Teruya-Feldstein J, Cordon-Cardo C, Lowe SW, Benezra R: **Mad2 overexpression promotes aneuploidy and tumorigenesis in mice.** *Cancer Cell* 2007, **11**:9-23.
28. Weigman VJ, Chao HH, Shabalin AA, He X, Parker JS, Nordgard SH, Grushko T, Huo D, Nwachukwu C, Nobel A, et al: **Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival.** *Breast Cancer Res Treat* 2012, **133**:865-880.
29. Agoulnik IU, Hodgson MC, Bowden WA, Ittmann MM: **INPP4B: the new kid on the PI3K block.** *Oncotarget* 2011, **2**:321-328.
30. Jackson S, Xiong Y: **CRL4s: the CUL4-RING E3 ubiquitin ligases.** *Trends Biochem Sci* 2009, **34**:562-570.
31. Sharma P, Nag A: **CUL4A ubiquitin ligase: a promising drug target for cancer and other human diseases.** *Open Biol* 2014, **4**:130217.
32. Chen LC, Manjeshwar S, Lu Y, Moore D, Ljung BM, Kuo WL, Dairkee SH, Wernick M, Collins C, Smith HS: **The human homologue for the *Caenorhabditis elegans* cul-4 gene is amplified and overexpressed in primary breast cancers.** *Cancer Res* 1998, **58**:3677-3683.
33. Gupta A, Yang LX, Chen L: **Study of the G2/M cell cycle checkpoint in irradiated mammary epithelial cells overexpressing Cul-4A gene.** *Int J Radiat Oncol Biol Phys* 2002, **52**:822-830.
34. Saucedo-Cuevas LP, Ruppen I, Ximenez-Embun P, Domingo S, Gayarre J, Munoz J, Silva JM, Garcia MJ, Benitez J: **CUL4A contributes to the biology of basal-like breast tumors through modulation of cell growth and antitumor immune response.** *Oncotarget* 2014, **5**:2330-2343.
35. Gastaldi S, Comoglio PM, Trusolino L: **The Met oncogene and basal-like breast cancer: another culprit to watch out for?** *Breast Cancer Res* 2010, **12**:208.

36. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, Felip E, Cappuzzo F, Paolini J, Usari T, et al: **First-line crizotinib versus chemotherapy in ALK-positive lung cancer.** *N Engl J Med* 2014, **371**:2167-2177.
37. Shaw AT, Yasothan U, Kirkpatrick P: **Crizotinib.** *Nat Rev Drug Discov* 2011, **10**:897-898.
38. Curigliano G, Goldhirsch A: **The triple-negative subtype: new ideas for the poorest prognosis breast cancer.** *J Natl Cancer Inst Monogr* 2011, **2011**:108-110.
39. Curigliano G: **New drugs for breast cancer subtypes: Targeting driver pathways to overcome resistance.** *Cancer Treat Rev* 2011, **38**:303-310.
40. Kannan N, Huda N, Tu L, Droumeva R, Aubert G, Chavez E, Brinkman RR, Lansdorp P, Emerman J, Abe S, et al: **The luminal progenitor compartment of the normal human mammary gland constitutes a unique site of telomere dysfunction.** *Stem Cell Reports* 2013, **1**:28-37.
41. Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nat Rev Drug Discov* 2004, **3**:711-715.
42. Scannell JW, Blanckley A, Boldon H, Warrington B: **Diagnosing the decline in pharmaceutical R&D efficiency.** *Nat Rev Drug Discov* 2012, **11**:191-200.
43. 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.

CHAPTER 5: THE MMTV-WNT1 MURINE MODEL PRODUCES TWO PHENOTYPICALLY DISTINCT SUBTYPES OF MAMMARY TUMORS WITH UNIQUE CLINICAL OUTCOMES TO EGFR INHIBITORS

OVERVIEW

Background

The Wnt gene family is an evolutionarily conserved group of proteins that regulate cell growth, differentiation, and stem cell self-renewal. Aberrant Wnt signaling in human breast tumors has been proposed to be an attractive drug target, especially in the basal-like subtype where canonical Wnt signaling is both enriched and predictive of poor clinical outcomes. The development of effective Wnt based therapeutics, however, has been slowed in part by a limited understanding of the context dependent nature with which these aberrations influence breast tumorigenesis.

Methods

MMTV-Wnt1 mice are an established model for studying Wnt signaling in breast tumors. We recently reported that this model develops two subtypes of tumors by gene expression classification: Wnt1-Early^{Ex} and Wnt1-Late^{Ex}. Here, we validate this initial observation using a combination of histology, fluorescence-activated cell sorting (FACS), limiting dilution assays, mutation analysis, gene expression profiling, and drug treatments to compare the phenotypes of these two Wnt1 tumor subtypes.

Results

Wnt1-Early^{Ex} tumors have high expression of canonical Wnt, non-canonical Wnt, and EGFR signaling pathway signatures. Therapeutically, Wnt1-Early^{Ex} tumors had a dynamic reduction in tumor volume when treated with EGFR inhibitors. Wnt1-Early^{Ex} tumors had primarily Cd49f⁺/Epcam⁻ FACS profiles, but were unable to be serially transplanted into wild-type FVB female mice. Wnt1-Late^{Ex} tumors, conversely, had a bloody gross pathology, which is highlighted by the presence of ‘blood lakes’ by H&E staining. These tumors had primarily Cd49f⁺/Epcam⁺ FACS profiles, but also contained a secondary Cd49f⁺/Epcam⁻ subpopulation. Both Wnt1-Late^{Ex} FACS subpopulations contained activating *Hras* mutations and were capable of individually reproducing tumors when serially transplanted into wild-type FVB female mice.

Conclusions

This study definitely shows that the MMTV-Wnt1 mouse model produces two phenotypically distinct subtypes of mammary tumors. Importantly, these subtypes differ in their therapeutic response to EGFR inhibitors, suggesting that a subset of human tumors with aberrant Wnt signaling may also respond to these drugs.

BACKGROUND

The mammalian breast is a unique organ capable of dynamic morphologic and physiologic change during organogenesis, puberty, pregnancy, lactation, and involution [1]. These processes are supported by a breast morphology that can be subdivided into four primary compartments: the stroma, the basement membrane, the basal layer, and the luminal layer [2, 3]. Within each of these compartments reside specific cell types that together form a mammary cell

hierarchy [4-6]. Specifically, the stroma consists primarily of fibroblasts, adipocytes, and immune cells [3, 7]. The basal layer is enriched for myoepithelial cells and mammary stem cells (MaSC) [8] and the luminal layer contains a combination of estrogen receptor (ER) positive and ER negative mature luminal cells [3].

Each cell within this hierarchy has developed specialized functions to support the necessary changes that will occur over a woman's lifetime. These processes include important elements of paracrine signaling to transmit signals across the different mammary compartments to specific recipients [9]. The Wnt family is an evolutionarily conserved group of proteins that promote paracrine signal transduction through at least five different pathways [2]. The canonical Wnt pathway signals through Frizzled (Fzd) and LDL-receptor-related (Lrp) co-receptors (Lrp5 and Lrp6) to activate Beta-catenin transcriptional regulation of key genes [2], such as c-Myc [10], c-Jun [11], and Vegf [12]. The other Wnt-regulated pathways are collectively referred to as non-canonical Wnt signaling. These include calcium and planar cell polarity signaling through Fzd receptors, Jnk signaling through the Ror2 receptor and Src signaling through the Ryk receptor [2]. While these pathways are commonly described from a cell autonomous perspective, complex signaling patterns emerge when paracrine signaling is considered [9]. In addition, there are 19 Wnt ligands and 10 Fzd receptors. When taking into account co-receptors and cell type specific expression patterns [13, 14], a large number of combinations are possible. Given the importance of Wnt signaling for controlling cell growth, differentiation, and stem cell self-renewal [15], a research emphasis has been placed on better understanding these Wnt signaling pathways.

One area of particular focus has been determining how aberrant Wnt signaling influences breast tumor formation and progression. Breast cancer is a heterogeneous disease that can be

segregated into at least six distinct intrinsic subtypes based on gene expression profiles: basal-like, claudin-low, HER2-enriched, luminal A, luminal B, and normal-like [16-18]. Interestingly, canonical Wnt signaling is enriched in basal-like breast tumors [19]. These patients also tend to have a poor clinical outcome [19], suggesting Wnt signaling as a potential therapeutic target [15]. Unlike colorectal cancer where inappropriate Wnt pathway activation is associated with gene mutations [20], mutations affecting Wnt associated genes are uncommon in breast tumors [18]. Instead, activation in breast tumors is proposed to occur through the downregulation of negative Wnt pathway regulators, such as secreted frizzles-related proteins [21, 22]. Even though these pathways have been highly examined, more research is needed to fully untangle the complex behavior of these signaling molecules. For instance, the molecular mechanisms that explain how paracrine Wnt signaling can induce growth of some tumors and inhibit it in others have remained elusive [23].

Genetically engineered mouse models are a useful resource for studying mammary tumors *in vivo* under genetically controlled and immune competent conditions [24]. MMTV-Wnt1 mice are particularly useful for modeling Wnt signaling in breast tumors [25, 26]. These tumors are comprised of mixed-lineage subclonal populations, having features of both luminal and basal epithelial cells [27]. In a subset of MMTV-Wnt1 tumors, cooperation between both subclonal populations is required for tumor propagation [27], highlighting this model as a tool for studying Wnt paracrine signaling and intratumoral heterogeneity [28]. We recently reported that MMTV-Wnt1 mice develop two subtypes/classes of tumors based on gene expression profiling [29], a finding that is surprisingly underrepresented in the vast literature on this model. Here, we investigate the significance of our prior observation and show that that these two classes of tumors have distinct phenotypes. Importantly, drug treatment experiments confirmed

that these subtypes had substantial response differences, thus shedding additional light on the significance of the Wnt pathway.

RESULTS

Even though the Wnt family has been highly studied from both developmental and oncology perspectives, the complexity of this pathway has hindered a complete understanding of the molecular mechanisms that regulate cell growth, differentiation, and stem cell self-renewal [15]. The MMTV-Wnt1 murine model is attractive for studying aberrant Wnt signaling in breast carcinoma [25, 26]. Interestingly, we find that these mice have a broad distribution of tumor latencies, developing as early as 5 weeks of age and as late as 58 weeks of age (Figure 20A). A histogram of 172 tumor latencies produces a bimodal distribution, with an ‘early’ local maximum around 7.5 weeks and a ‘late’ local maximum around 21.5 weeks. Even though no differences were observed on a DNA copy number level (Figure 21), gene expression profiling found that these Wnt-Early^{Ex} and Wnt1-Late^{Ex} tumors have distinct biological features [29], indicating that this tumor latency distribution is more than just a stochastic event (Figure 20B). Given these findings, we performed an in depth comparison of these two Wnt1 tumor classes to further our understanding of the clinical significance of Wnt signaling in breast carcinoma.

Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors have distinct gross pathology and histology traits

In addition to being classified into different molecular expression subtypes/classes, Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors were also found to have distinct gross pathological features. Specifically, Wnt1-Early^{Ex} tumors tended to show a dense cellular morphology and be

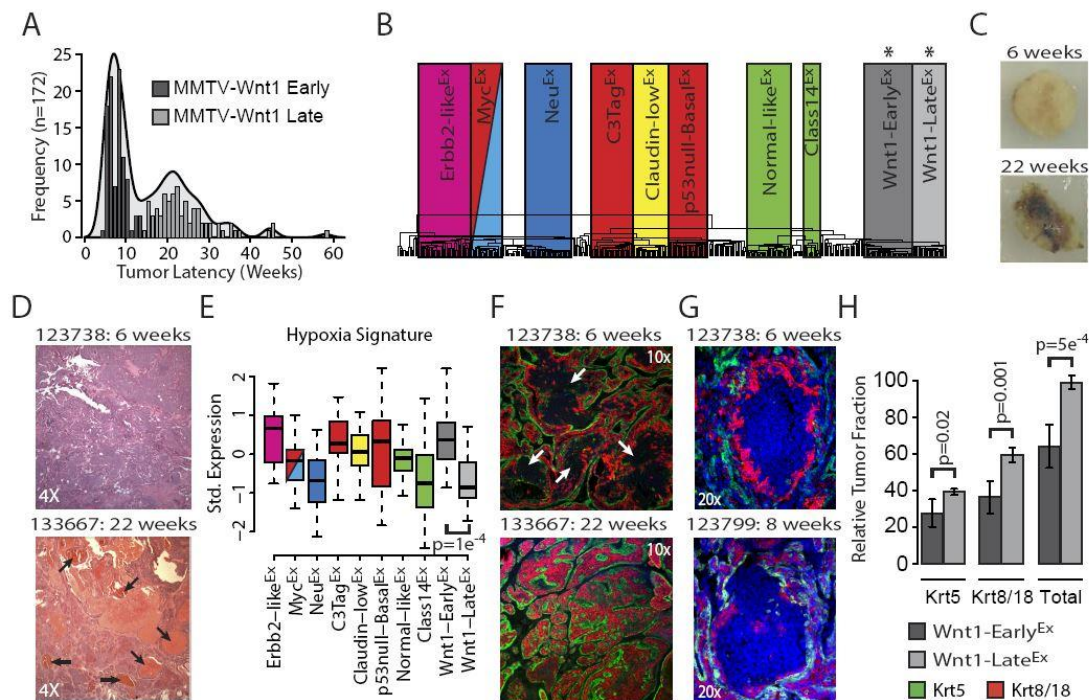


Figure 20: The MMTV-Wnt1 model produce two classes of mammary tumors with distinct latencies, gross pathology and histology features

A. MMTV-Wnt1 tumor latency histogram by week **B.** Dendrogram of a hierarchical cluster of all murine tumors in our dataset using a previously defined intrinsic gene list [29]. Boxes correspond to previously defined murine intrinsic subtypes/classes [29]. **C.** Gross pathology of representative Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors. **D.** H&E staining of representative Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors. **E.** Standardized expression of a hypoxia gene signature [30] across mouse class. **F.** Krt5 and Krt8/18 staining of representative Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors. **G.** Krt5 and Krt8/18 staining of representative Wnt1-Early^{Ex} tumors. **H.** Relative tumor fraction of Krt5 positive and Krt8/18 positive cells.

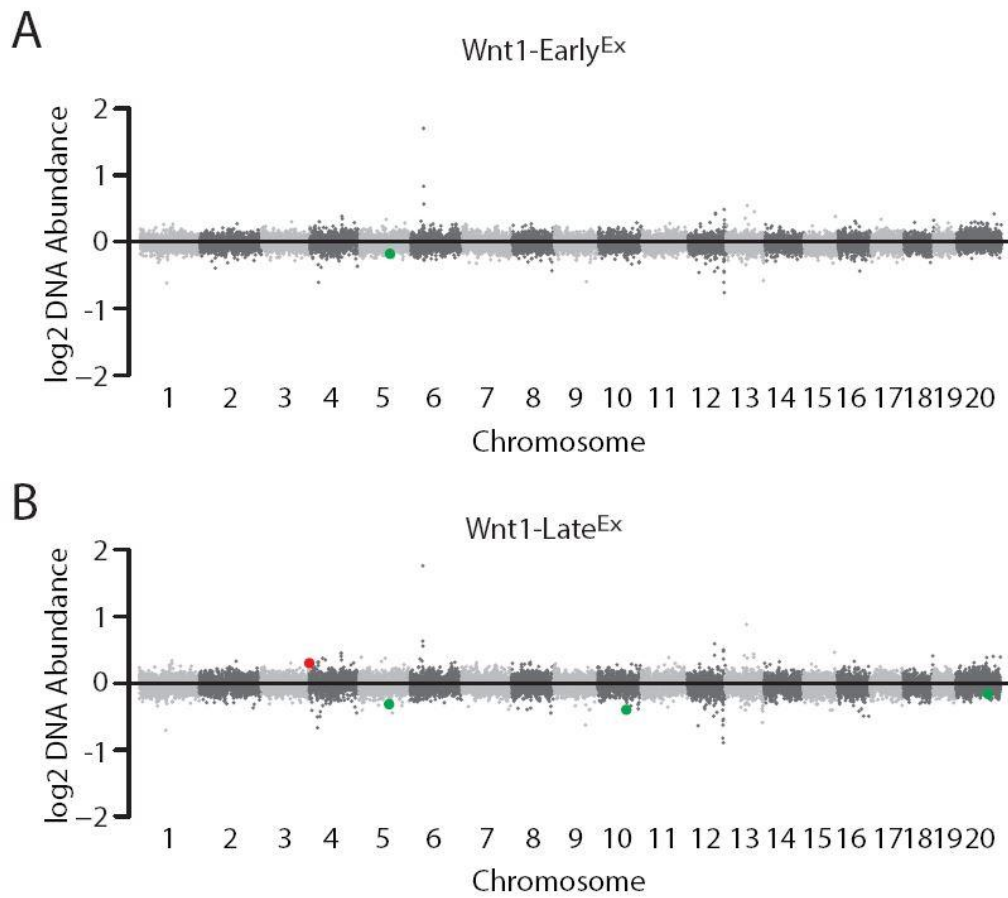


Figure 21: Wnt1-Early^{Ex} and Wnt1-Late^{Ex} have similar DNA copy number landscapes
 Displayed in genomic order are the median class DNA copy number levels for **A.** Wnt1-Early^{Ex} and **B.** Wnt1-Late^{Ex} tumors.

more resistant to incision, while Wnt1-Late^{Ex} tumors appeared to be more necrotic and often filled with pockets of blood (Figure 20C). Importantly, these characteristics were found to be irrespective of tumor size at the time of collection, indicating that these observations were not a technical artifact but inherent to the tumors themselves. These gross pathological differences were recapitulated with hematoxylin and eosin (H&E) staining in which Wnt1-Late^{Ex} tumors had characteristic ‘blood lake’ regions as highlighted by black arrows in Figure 20D. These different vascular traits led us to hypothesize that Wnt1-Early^{Ex} tumors should be more hypoxic than Wnt1-Late^{Ex} tumors. In support of this, a VEGF/Hypoxia gene signature [30] is highly expressed in Wnt1-Early^{Ex} tumors and lower expressed in Wnt1-Late^{Ex} tumors.

It is well documented that MMTV-Wnt1 tumors are comprised of mixed-lineage subclonal populations, having features of both luminal and basal cells [27]. To investigate the relative fraction of these subclonal populations, immunofluorescence staining was performed using antibodies against Krt5 (a marker of basal cells) and Krt8/18 (a marker of luminal cells). Consistent with the literature, both Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors stained positive for both cell populations (Figure 20F). Unlike Wnt1-Late^{Ex}, Wnt1-Early^{Ex} tumors had distinct regions that did not stain positive for either Krt5 or Krt8/18 (Figure 20G). These areas did stain positive for DAPI, indicating that there are cells within these regions. When measuring the relative tumor fractions of these populations, it is observed that Wnt1-Late^{Ex} tumors contained a higher fraction of Krt5 positive cells ($p=0.02$) and Krt8/18 positive cells ($p=0.001$) than Wnt1-Early^{Ex} tumors (Figure 20H). Combined, these two cell fractions comprised about 100% of Wnt1-Late^{Ex} tumors. For Wnt1-Early^{Ex} tumors, these two fractions only accounted for only about 65% of the tumor, with the remaining 35% consisting of regions that did not stain positive for either Krt5 or Krt8/18. Further experiments, such as laser capture microdissection (LCM), will

be needed to determine the biological significance of these heterogeneously staining (or not staining) tumor regions.

Wnt1-Early^{Ex} tumors are enriched for canonical and non-canonical Wnt pathway signatures

Wnt signal transduction can occur through several different molecular pathways, including canonical Wnt, Jnk, and Src [2]. To investigate these pathways in our mouse tumors, expression based pathway gene signatures were used to estimate pathway activity [29, 31]. As a good positive control, the canonical KEGG WNT signaling pathway was the most highly expressed in both Wnt1-Early^{Ex} and Wnt1-Late^{Ex} classes (Figure 22A). Interestingly, this pathway signature is higher expressed in Wnt1-Early^{Ex} compared to Wnt1-Late^{Ex} tumors ($p=0.02$). This observation does not appear to be a result of variation in Wnt1 transgene expression, as there was no statistical difference between the Wnt1-Early^{Ex} and Wnt1-Late^{Ex} classes for the single Wnt1 probe (exon 4) in our combined murine dataset (Figure 22B). A closer investigation into the individual genes within the canonical KEGG WNT signaling pathway identified several differentially expressed genes between Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors (Figure 22B). Canonical Wnt signaling occurs through Frizzled (Fzd) and LDL-receptor-related (Lrp) co-receptors [2]. Fzd receptors *Fzd1*, *Fzd2*, *Fzd9*, and *Fzd10* are higher expressed in Wnt1-Early^{Ex} tumors, while *Fzd5* is higher in Wnt1-Late^{Ex} tumors (FDR 0%). In addition, the *Lef1* transcription factor and its target, *c-Jun* [11], are also higher expressed in Wnt1-Early^{Ex} tumors (FDR 0%). These results are consistent with higher canonical Wnt pathway activity in Wnt1-Early^{Ex} tumors.

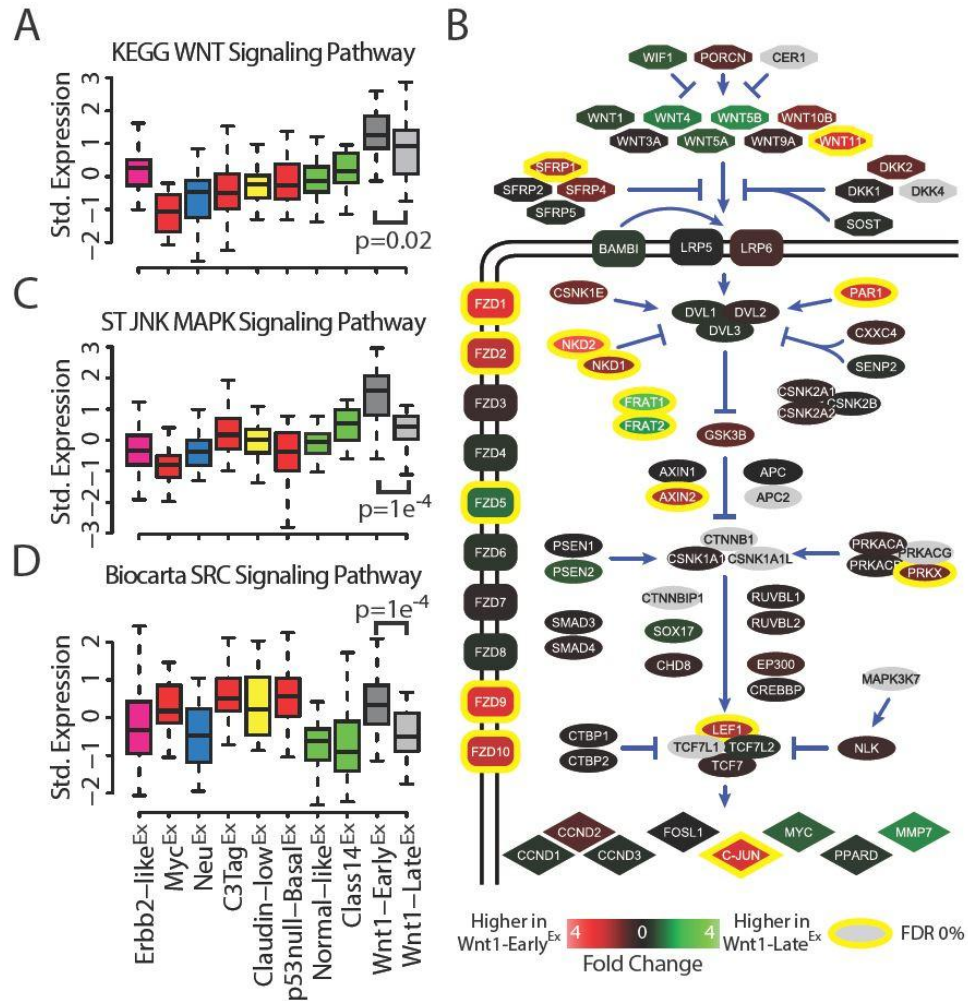


Figure 22: Wnt-Early^{Ex} tumors have expression of Wnt associated pathway signatures
A. Standardized expression of the KEGG Wnt signaling pathway signature across mouse class.
B. Schematic of the KEGG Wnt signaling pathway. **C.** Standardized expression of the ST JNK MAPK signaling pathway signature across mouse class. **D.** Standardized expression of the Biocarta SRC signaling pathway signature across mouse class.

In addition to higher expression of the canonical KEGG WNT signaling pathway signature, Wnt1-Early^{Ex} tumors also have higher expression of non-canonical Wnt signaling pathway signatures: ST JNK MAPK Signaling Pathway ($p=1.0e^{-4}$) (Figure 22C) and Biocarta SRC Signaling Pathway ($p=1.0e^{-4}$) (Figure 22D). These results are intriguing because they suggest that Wnt1-Early^{Ex} tumors signal through both canonical and non-canonical Wnt pathways to a greater extent than Wnt1-Late^{Ex} tumors.

Wnt1-Early^{Ex} tumors respond to epidermal growth factor receptor inhibitors

In addition to canonical and non-canonical signaling, Wnt associated genes can also crosstalk with a variety of other signal transduction pathways [32-34], including epidermal growth factor receptor (EGFR) signaling [35]. Specifically, NKD2 is capable of binding and shuttling TGFA to the plasma membrane, which serves as an activating ligand of EGFR (Figure 23A). Interestingly, both *Nkd2* and *Tgfa* are higher expressed in Wnt1-Early^{Ex} tumors (FDR 0%). Given this observation, we hypothesized that Wnt1-Early^{Ex} tumors might have a greater degree of EGFR signaling than Wnt1-Late^{Ex} tumors. Consistent with this, the KEGG EGFR signaling pathway signature is higher expressed in Wnt1-Early^{Ex} as compared to Wnt1-Late^{Ex} tumors ($p=0.001$) (Figure 23B). To determine the clinical importance of these findings, Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors were randomized into one of three treatment groups: untreated, erlotinib (an EGFR inhibitor), or lapatinib (a dual EGFR and HER2 inhibitor). As hypothesized, Wnt1-Early^{Ex} tumors had a median tumor regression of 90% when treated with erlotinib and a median tumor regression of 85% when treated with lapatinib at the end of the two week treatment period (Figure 23C). Wnt1-Late^{Ex} tumors, however, continued to progress with erlotinib treatment with a median tumor growth of 109%. When treated with lapatinib, Wnt1-Late^{Ex} tumors had a

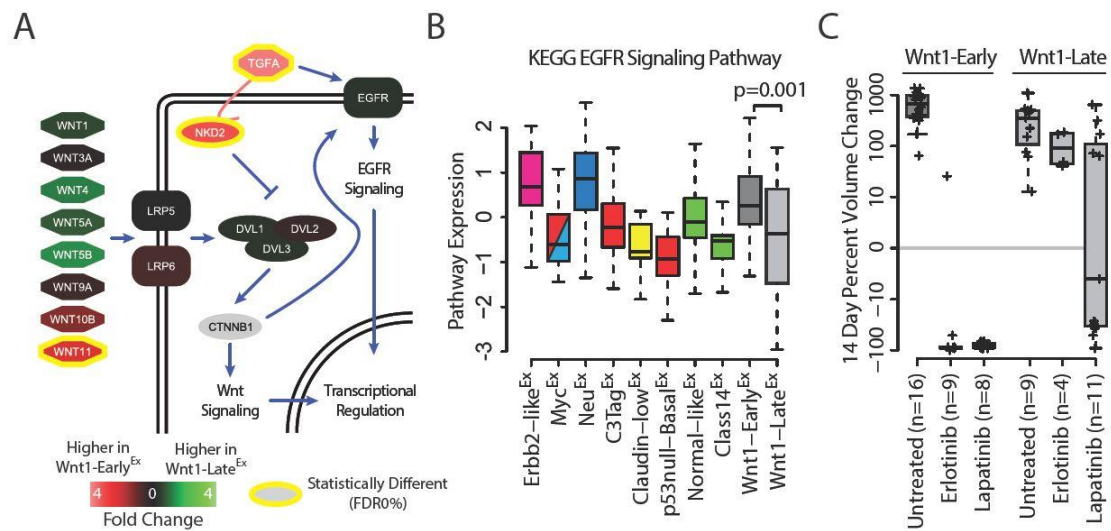


Figure 23: Wnt-Early^{Ex} tumors respond to EGFR inhibitors

A. Schematic of WNT and EGFR pathway crosstalk **B.** Standardized expression of the KEGG EGFR signaling pathway signature across mouse class. **C.** 14 day tumor response to Erlotinib and Lapatinib treatment.

bimodal response with about half of tumors regressing with treatment and half progressing.

These results indicate that Wnt1-Early^{Ex} tumors were therapeutically more responsive to EGFR inhibitors than Wnt1-Late^{Ex} tumors.

Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors have distinct mammary subpopulation FACS profiles

Normal mammary gland physiology is supported by an underlying, complex cell hierarchy [4-6]. A simplistic model places the multi-potent mammary stem cell (MaSC) at the base of this hierarchy, having extensive, self-regenerative potential [3]. During mammary development, the MaSC has been proposed to divide asymmetrically to produce basal/myoepithelial cells as well as luminal progenitors (LumProg), which have more restricted proliferative and differentiation capabilities [3]. LumProg cells are capable of further differentiation into mature luminal (MatureLum) cells, such as Estrogen Receptor (ER)-positive ductal epithelium, which have an even more limited proliferative potential and some of which are terminally differentiated [3].

MMTV-Wnt1 tumors may originate from several, if not all, of the cell types within this mammary hierarchy. To determine if Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors share features with any of these cell populations, four primary tumors from each class were fluorescence-activated cell sorted (FACS) using antibodies against Cd49f and Epcam [27, 36]. FACS profiles of Wnt1-Early^{Ex} tumors consisted of two populations (Figure 24A). The major (~75%) epithelial cell population was Cd49f⁺/Epcam⁻, while the minor (~10%) population was Cd49f⁺/Epcam⁺ (Figure 24B). Normal human MaSCs are defined as having CD49f⁺/EpCAM⁻ FACS profiles [3], indicating that the majority of Wnt1-Early^{Ex} tumor cells share similar features as normal MaSCs.

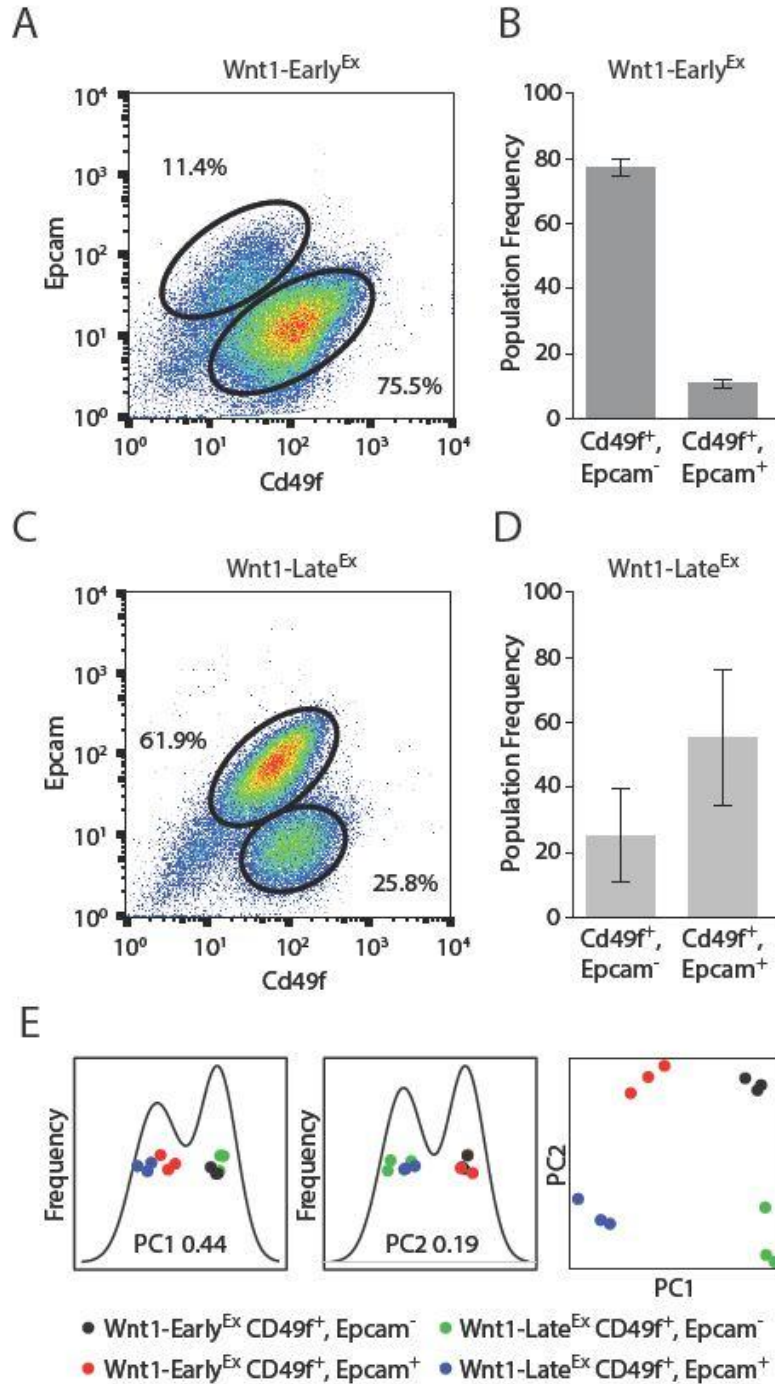


Figure 24: Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors share features with different normal mammary cell types

A. Cd49f, Epcam FACS profile of a representative Wnt1-Early^{Ex} tumor **B.** FACS population frequencies of Wnt1-Early^{Ex} tumors **C.** Cd49f, Epcam FACS profile of a representative Wnt1-Late^{Ex} tumor **D.** FACS population frequencies of Wnt1-Late^{Ex} tumors **E.** First two principle components of FACS sorted Wnt1 tumors.

Although Wnt1-Late^{Ex} tumor FACS profiles also had two FACS populations, the frequencies were distinct from Wnt1-Early^{Ex} tumors (Figure 24C). Specifically, the major (~60%) epithelial cell population was Cd49f⁺/Epcam⁺, while the minor population (~25%) was Cd49f⁺/Epcam⁻ (Figure 24D). Normal human LumProg cells are defined as having CD49f⁺/EpCAM⁺ FACS profiles [3], indicating that the majority of Wnt1-Late^{Ex} tumor cells share similar features as normal LumProg cells.

It is possible that the MMTV-Wnt1 model produces semi-homogeneous tumors [29] simply because of intratumor variation in the frequencies of these two FACS populations and not because of differences between corresponding FACS populations across the two classes themselves. For example, this hypothesis would propose that the Wnt1-Early^{Ex} Cd49f⁺/Epcam⁻ population should be phenotypically the same as the Wnt1-Late^{Ex} Cd49f⁺/Epcam⁻ population. To address this, three tumors from each Wnt1 class were FACS into their corresponding populations and microarray analyzed. A global transcriptomic comparison of these FACS populations using a principle component (PC) analysis highlights that the first PC separates the Cd49f⁺/Epcam⁻ population from the Cd49f⁺/Epcam⁺ population, irrespective of which Wnt1 tumor class they were derived from (Figure 24E). This observation is consistent with the proposed hypothesis that these FACS populations are phenotypically similar across classes, but the first PC only explains 44% of the variation. The second PC, which explains 19% of the variation, separates Wnt1-Early^{Ex} from Wnt1-Late^{Ex} tumors. Taken together, these results indicate that while the corresponding FACS populations are highly similar across Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors, they also have class specific features.

Both Wnt1-Late^{Ex} tumor FACS subpopulations have tumor initiating potential

Given that Wnt1-Early^{Ex} tumors share features of normal MaSCs and Wnt1-Late^{Ex} tumors share features of normal LumProg cells, we hypothesized that these two Wnt1 classes may have different tumor initiating potential. To test this, Wnt1 tumors were FACS sorted into their subpopulations and a limiting dilution assay was performed in which each subpopulation was injected into the mammary pad of female FVB wild-type mice. In addition, a subset of MMTV-Wnt1 tumors requires both FACS populations for tumor growth [27]; therefore, a third cohort consisting of an equal mixture of each FACS subpopulation was also performed to investigate this possibility in our two classes. Interestingly, Wnt1-Early^{Ex} tumor cells were unable to be serially transplanted into wild-type mice, giving rise to no tumors after injection with 50,000 cells (Figure 25). Conversely, all combinations of Wnt1-Late^{Ex} tumor cells gave rise to tumors. These results were unexpected, but indicate that both Wnt1-Late^{Ex} FACS subpopulations have tumor initiating potential.

Tumors that arose from the individual Wnt1-Late^{Ex} FACS populations were then re-FACS analyzed to investigate their tumor profiles. Similar to the parental tumor, the FACS profile of Wnt1-Late^{Ex} Cd49f⁺/Epcam⁻ injected cells contained two populations (Figure 25B) of about equal frequency (Figure 25C). A similar observation was observed for Wnt1-Late^{Ex} Cd49f⁺/Epcam⁺ injected cells (Figure 25D and 5E). These results show that both Wnt1-Late^{Ex} populations are capable of reproducing the other population when injected into the mammary pad after FACS purification.

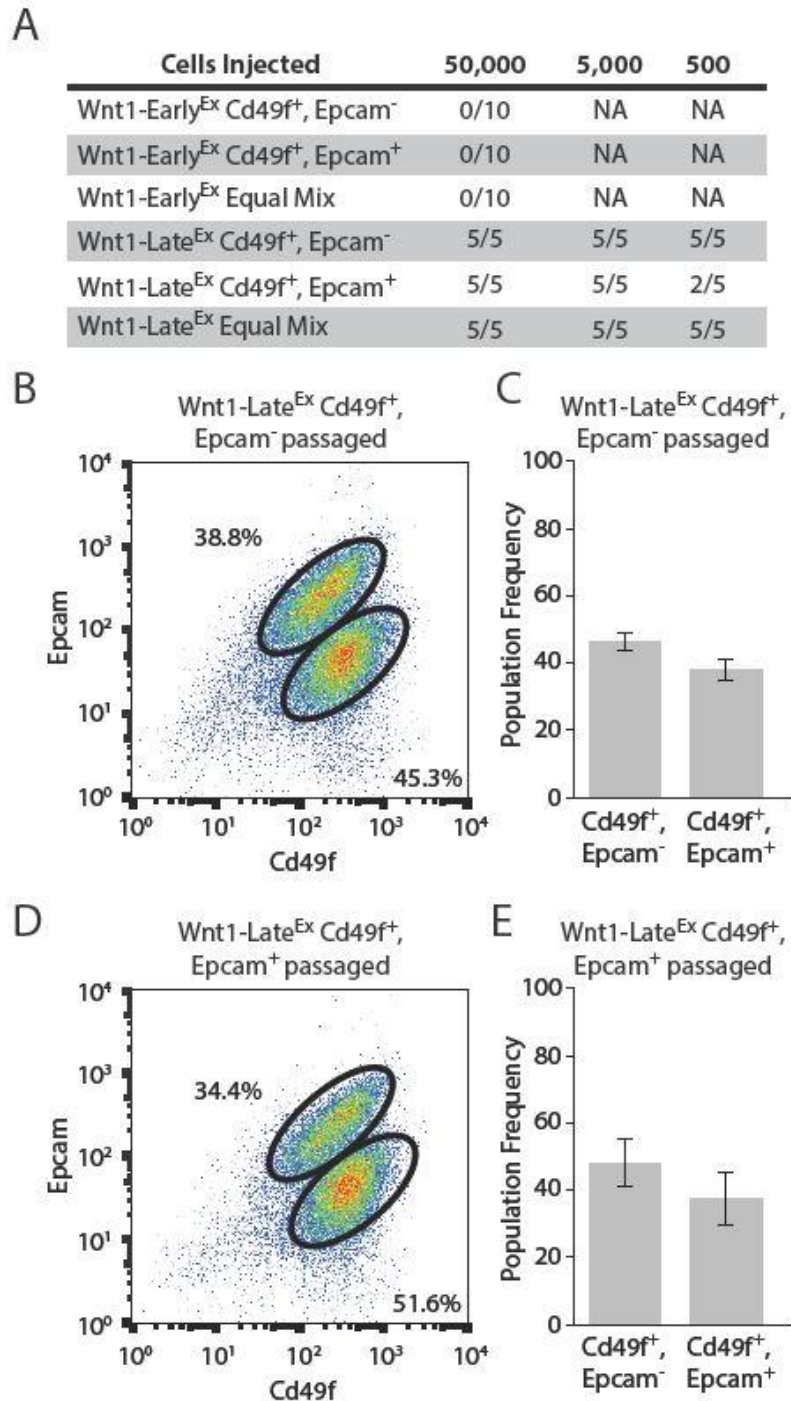


Figure 25: Both Wnt1-Late^{Ex} tumor subpopulations have tumor initiating potential

A. Limiting dilution cell transplantation assay **B.** Cd49f, Epcam FACS profile of a representative Wnt1-Late^{Ex} CD49f⁺, Epcam⁻ passaged tumor. **C.** FACS population frequencies of Wnt1-Late^{Ex} CD49f⁺, Epcam⁻ passaged tumors **D.** Cd49f, Epcam FACS profile of a representative Wnt1-Late^{Ex} CD49f⁺, Epcam⁺ passaged tumor. **E.** FACS population frequencies of Wnt1-Late^{Ex} CD49f⁺, Epcam⁺ passaged tumors.

Both Wnt1-Late^{Ex} tumor FACS subpopulations have activating *Hras1* mutations

To highlight the possible genetic drivers which explain this difference in tumor initiating potential, five Wnt1-Early^{Ex} and five Wnt1-Late^{Ex} tumors were RNA sequenced (i.e. mRNA-seq) to profile mutations in MMTV-Wnt1 tumors. It is known that a subset of MMTV-Wnt1 tumors harbor activating *Hras1* mutations [27]. Interestingly, all of the Wnt1-Late^{Ex} tumors profiled contained exon 3 activating *Hras1* mutations, while none of the Wnt1-Early^{Ex} tumors were mutated (Figure 26A). Even though the specific mutations varied, all of the Wnt1-Late^{Ex} *Hras1* mutations resulted in an amino acid substitution at position 61, and the majority of samples had a mutation allele frequency of ~25% (Figure 26B).

In colorectal cancer, *APC* loss of function mutations synergize with *KRAS* activating mutations to activate cancer stem cells [37]. If similar synergy occurs in Wnt1-Late^{Ex} tumors, these activating *Hras1* mutations may explain why these tumors have tumor initiating potential. This hypothesis would predict that both Wnt1-Late^{Ex} FACS populations should contain *Hras1* mutations since both were capable of producing tumors when injected individually. To test this, DNA was extracted from each FACS population and Sanger sequenced. In support of this hypothesis, both Wnt1-Late^{Ex} populations contained *Hras1* mutations, while both of the Wnt1-Early^{Ex} populations were wild-type. The initial *Hras1* mutations were identified from RNAseq data; therefore, it is possible that Wnt1-Early^{Ex} tumors could contain DNA *Hras1* mutations but are transcriptionally repressing that allele. The Sanger sequencing data indicates that Wnt1-Early^{Ex} tumors do not contain DNA *Hras1* mutations.

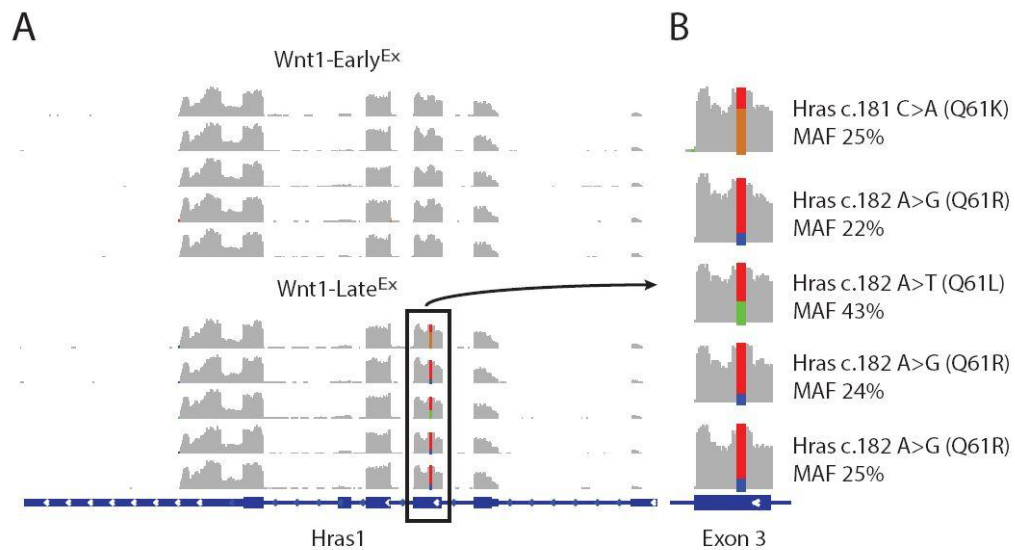


Figure 26: Wnt-Late^{Ex} tumors have activating Hras mutations

A. *Hras1* RNAseq reads. Gray read depth corresponds to wild-type sequence. **B.** *Hras1* exon 3 mutations in Wnt1-Late^{Ex} tumors.

DISCUSSION

Breast cancer is the second leading cause of cancer related deaths in American women [38]. Patient care is particularly complicated by breast tumor heterogeneity, which is defined by multiple intrinsic subtypes [16-18]. While a greater understanding of tumor biology has led to targeted treatment options for most of these subtypes [39, 40], personalized drug targets for basal-like tumors remain an important unmet clinical need [41]. Given that Wnt signaling is both enriched in basal-like tumors and predictive of poor clinical outcomes [19], Wnt signaling has been proposed to be an attractive drug target for these patients [15]. The development of effective Wnt based therapeutics against breast tumors, however, has been slowed in part by the complexity of Wnt signaling [2, 15]. In addition to canonical and several non-canonical pathways, Wnt signaling is also known to crosstalk with a variety of other signal transduction pathways [32-35]. It is this context dependent nature that likely accounts for the finding that paracrine Wnt signaling can induce growth of some tumors and inhibit it in others [23]. Given the importance of Wnt signaling for regulating cell growth differentiation, and stem cell self-renewal [15], a better understanding of these signaling pathways is needed.

MMTV-Wnt1 mice are an attractive model for studying the context dependent nature with which these aberrations influence breast tumorigenesis [27]. We recently reported that this model develops two subtypes of tumors by gene expression classification [29]. Given that this finding is underrepresented in the literature, we sought to validate our initial observation with a more thorough examination of these two Wnt1 subtypes. Here we definitely show that these two tumor subtypes are indeed phenotypically and clinically distinct, furthering our understanding of Wnt signaling in breast cancer.

Wnt1-Early^{Ex} mice were initially characterized by their early tumor latency, accounting for ~60% of the MMTV-Wnt1 tumors profiled in this study. In addition to having a cellular dense gross pathology, these tumors were also enriched for a hypoxia gene signature. While these tumors contained features of both basal and luminal mammary cell types, they also have distinct regions that do not stain positive for either cell marker. Although the results presented here are unable to address the biological impact of these regions, we propose that they are likely to be significant given that they account for ~35% of the tumor. On a pathway level, Wnt1-Early^{Ex} tumors have higher expression of both canonical and non-canonical signaling. On a gene level, *Nkd2* and *Tgfa* were particularly highly expressed in these tumors and were capable of crosstalk with EGFR signaling [35]. We validated the clinical significance of this observation by treating MMTV-Wnt1 tumors with Erlotinib and Lapatinib. As predicted, Wnt1-Early^{Ex} tumors have a dynamic reduction in tumor volume after only 14 days of treatment. The FACS profile of these tumors highlights that they are ~75% Cd49f⁺/Epcam⁻. Although this profile is similar to adult mammary stem cells (MaSCs) [3], these tumors are unable to be serially transplanted into wild-type FVB female mice.

Wnt1-Late^{Ex} mice were initially identified by their longer tumor latency, accounting for ~40% of the MMTV-Wnt1 tumors profiled in this study. These tumors have a bloody gross pathology, which is highlighted by the presence of ‘blood lakes’ by H&E staining. They also have cell features of both basal and luminal mammary cell types, which account for ~100% of the tumor. On a pathway level, Wnt1-Late^{Ex} tumors have high expression of canonical Wnt signaling compared to other mouse tumors, but lower expression in comparison to Wnt1-Early^{Ex} tumors. Although these tumors have a FACS profile that is ~60% Cd49f⁺/Epcam⁺, they also contain a secondary Cd49f⁺/Epcam⁻ population that accounts for ~25% of epithelial cells.

Interestingly, both of these cell populations are able to reproduce tumors when serially transplanted into wild-type FVB female mice. Similar to the parental tumor, the FACS profile of these serially transplanted cells contains two populations, indicating that both Wnt1-Late^{Ex} populations are capable to reproducing the other. This tumor initiating property may be linked to the presence of *Hras1* mutations in Wnt1-Late^{Ex} tumors. This hypothesis is supported by the finding that *KRAS* mutations synergize with aberrant Wnt signaling in colorectal cancer to activate cancer stem cells [37]. In a previous publication on MMTV-Wnt1 tumors, *Hras1* mutations were shown to be specific to the Cd49f⁺/Epcam⁻ population [27], but in the tumors we sequenced, *Hras1* mutations were identified in both the Cd49f⁺/Epcam⁻ and the Cd49f⁺/Epcam⁺ FACS populations. This finding is also consistent with our hypothesis since we would predict that only cell populations with *Hras1* mutations are capable of being serially passaged. Further testing should be aimed at addressing this possibility by introducing *Hras* mutations into Wnt1-Early^{Ex} tumors and retesting their tumor initiating potential.

The FACS profiles imply that Wnt1-Early^{Ex} tumors arise from MaSCs and that Wnt1-Late^{Ex} tumors arise primarily from luminal progenitor cells. Although additional experiments (e.g. lineage tracing) will be required to unequivocally determine this, these associations at the very least identify which normal mammary subpopulation a given tumor most represents in its current state. If the inappropriate expansion of these cell populations was truly a stochastic event, we would not expect to observe such a stark contrast in tumor latency between the two Wnt1 subtypes. This latency difference suggests that these are not random, but regulated events. Of the tumors profiled in this study, we did not find a single case of subtype switching (where an ‘early’ latency tumor was classified as a ‘Wnt1-Late^{Ex}’ or vice versa), indicating that these regulating mechanisms are rather strong. Broadly, we hypothesize that the mechanisms governing Wnt1-

Early^{Ex} tumor development are likely related to puberty, which occurs around this time [6]. Wnt1-Late^{Ex} tumor susceptibility is likely influenced by age related changes that increase the risk of developing *Hras1* activating mutations, as is the case for *KRAS* mutations in colorectal cancer [42].

MATERIALS AND METHODS

Mouse husbandry

All animal work was done in UNC DLAM facilities in compliance with IACUC approved protocols. FVB/n mice carrying the MMTV-Wnt1 transgene were bred and housed until the onset of a mammary tumor. The following PCR primers were used for genotyping: 5'-GGACTTGCTTCTCTTCTCATAGCC-3' and 5'-CCACACAGGCATAGAGTGTCTGC-3'.

Gene expression

Microarray gene expression data from 27 murine models of mammary carcinoma and normal mammary tissue were downloaded from the following gene expression omnibus (GEO) entries: GSE3165, GSE8516, GSE9343, GSE14457, GSE15263, GSE17916, GSE27101, and GSE42640 [29]. An additional 35 MMTV-Wnt1 tumors were microarray profiled as previously described [29]. The 420 sample dataset was normalized to correct for microarray platform bias as previously described [29].

RNAseq libraries were prepared from 26 MMTV-Wnt1 tumors using a TruSeq RNA kit (Illumina #RS-122-2001) before being submitted to the Lineberger Comprehensive Cancer Center Genomics Core to be run on the Illumina HiSeq 2000.

Tumor differentiation scores were calculated across the microarray dataset as previously described [17]. Gene expression signatures were created for Wnt1-Early^{Ex} and Wnt1-Late^{Ex} tumors by performing a two-class (class X versus all others or Wnt1-Early^{Ex} vs Wnt1-Late^{Ex}) Significance Analysis of Microarrays (SAM) analysis on the microarray dataset [43]. Signatures were defined as all genes highly expressed in the class of interest with a false discovery rate (FDR) of 0%. Similarly, pathway signatures were created as previously described [29]. Expression scores for each gene and pathway signature were determined by calculating the standardized mean expression of the signature within each sample.

DNA copy number

Genomic DNA was extracted from 11 Wnt1-Early^{Ex} and 10 Wnt1-Late^{Ex} tumors using a DNeasy blood and tissue kit (Qiagen #69504), labeled with a Sure Tag DNA kit (Agilent #5190-4240), and hybridized to 244K CGH microarrays (Agilent #G4415A) as previously described [44].

The 21 sample aCGH dataset was extracted from the UNC Microarray Database as log₂ Cy5/Cy3 ratios, filtering for probes with Lowess normalized intensity values greater than ten in the control channel and for probes with data on greater than 70% of the microarrays [44]. The probes that passed these filters were then oriented in genomic order and a ten probe average was calculated on consecutive groups of ten probes across each chromosome, resulting in a final dataset of 23,204 features. A two-class (Wnt1-Early^{Ex} vs Wnt1-Late^{Ex}) SAM analysis was performed to identify genomic regions of amplification or deletion unique to each class (FDR of 0%).

Immunofluorescence

Tissue samples were fixed in 10% neutral buffered formalin (Sigma-Aldrich Cat#HT5012) overnight before being submitted to the UNC Lineberger Animal Histopathology core facility to be paraffin embedded and sectioned. Slides were heated for 30 minutes at 55°C and then washed with Xylene (Fisher Scientific Cat#X3P) and ethanol (Decon Laboratories, Inc Cat#2716) to deparaffin the samples. To increase antigen exposure, slides were boiled for 15 minutes in Antigen Retrieval Citra Plus Solution (BioGenex Cat#HK080-9K). Samples were blocked for 1 hour at room temperature in TBS (BioRad Cat#170-6435)/0.05% Tween 20 (BioRad Cat#161-0781) plus 5% normal goat serum (Sigma- Aldrich Cat#G9023). Proteins were labeled with murine Krt5 (Covance Cat#PRB-160P) and murine Krt8/18 (Fitzgerald Cat#20R-CP004) primary antibodies at 4°C overnight before being labeled with an anti-rabbit secondary antibody (Molecular Probes Cat#A11034) and anti-guinea pig (Molecular Probes Cat#A11076) at room temperature for one hour. Slides were mounted with DAPI (Vector Laboratories Cat#H-1500). Slides images were taken using a Nikon Eclipse E600 microscope and processed using the ImageJ software.

Drug Treatment

The Wnt1-Early^{Ex} tumor with the longest latency was 13.5 weeks and the Wnt1-Late^{Ex} tumor with the shortest latency was 16 weeks. From these observations, a 15 week cutoff was used to define Wnt1-Early and Wnt1-Late tumors for drug treatment analysis. MMTV-Wnt1 tumors were randomized into treatment groups and tumor growth was monitored using two-dimensional caliper measurements ($\text{Volume} = [(\text{width})^2 \times \text{length}] / 2$) [45]. Drug compounds were obtained from commercial sources (erlotinib from Genentech, Inc and lapatinib from

GlaxoSmithKline) before being synthesized into chow by OpenSource Diets to a final concentration of 25 mg/kg for erlotinib and 220 mg/kg for lapatinib [45]. Biological inhibitors were dosed continuously for two weeks. The percent change in tumor volume at the end of the 14 day treatment period was used to quantify response.

Flow cytometry

MMTV-Wnt1 tumors were dissociated into a single cell suspension using the following steps. First, each tumor was manually cut into small pieces with a razor blade in 1X collagenase/hyaluronidase (StemCell #07919) EpiCult media (StemCell #05601) before being placed in a rotator for two hours at 37°C. Following lysis of red blood cells using ammonia chloride (StemCell #07850), the tumors were incubated in 1X trypsin-EDTA (Sigma #T4049) for five minutes at 37°C and then in a 1X Dispase (StemCell #07923) DNase I solution (StemCell #07900) for five min at 37°C to reduce cell clumping. Cells were filtered through a 40µm nylon cell strainer (Fisher Scientific #08-771-1) in HBSS media (StemCell #37150) with 10% FBS (Sigma #F2442) to obtain the final single cell suspension. To remove non-epithelial cells, the single cell suspension was taken through a mouse epithelial cell enrichment kit (StemCell #19758) following the manufacturer's protocol. Cells were labeled with the following antibodies for 30 minutes at 4°C: FITC anti-mouse Epcam (eBioscience #11-5791-82) and APC anti-mouse Cd49f (eBioscience #17-0495-82). Fluorescence-activated cell sorting (FACS) was performed using a Beckman-Coulter CyAn ADP instrument and analyzed using the FlowJo v10 software program.

Sanger Sequencing

Genomic DNA was extracted from the six FACS purified fractions using a DNeasy blood and tissue kit (Qiagen #69504) and a portion of *Hras1* was PCR amplified using a Taq PCR kit (Qiagen Cat #201223) with the following primers: 5'-ATGGGGTATGATCCATCAGG-3' and 5'-CACACGGAACCTTCCTCAC-3' (Sigma-Aldrich). PCR products were enriched with a PCR purification Kit (Qiagen Cat#28104) before being submitted to the UNC Genome Analysis Facility for Sanger sequencing. Results were analyzed using the Sequencher software.

REFERENCES

1. Gjorevski N, Nelson CM: **Integrated morphodynamic signalling of the mammary gland.** *Nat Rev Mol Cell Biol* 2011, **12**:581-593.
2. Roarty K, Rosen JM: **Wnt and mammary stem cells: hormones cannot fly wingless.** *Curr Opin Pharmacol* 2010, **10**:643-649.
3. Visvader JE: **Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis.** *Genes Dev* 2009, **23**:2563-2577.
4. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C: **Distinct stem cells contribute to mammary gland development and maintenance.** *Nature* 2011, **479**:189-193.
5. Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, Harrell JC, McNamara G, Schwede M, Culhane AC, et al: **Taxonomy of breast cancer based on normal cell phenotype predicts outcome.** *J Clin Invest* 2014, **124**:859-870.
6. Visvader JE, Stingl J: **Mammary stem cells and the differentiation hierarchy: current status and perspectives.** *Genes Dev* 2014, **28**:1143-1158.
7. Arendt LM, Rudnick JA, Keller PJ, Kuperwasser C: **Stroma in breast development and disease.** *Semin Cell Dev Biol* 2010, **21**:11-18.
8. Woodward WA, Chen MS, Behbod F, Rosen JM: **On mammary stem cells.** *J Cell Sci* 2005, **118**:3585-3594.
9. Rosen JM, Roarty K: **Paracrine signaling in mammary gland development: what can we learn about intratumoral heterogeneity?** *Breast Cancer Res* 2014, **16**:202.
10. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW: **Identification of c-MYC as a target of the APC pathway.** *Science* 1998, **281**:1509-1512.
11. Mann B, Gelos M, Siedow A, Hanski ML, Gratchev A, Ilyas M, Bodmer WF, Moyer MP, Riecken EO, Buhr HJ, Hanski C: **Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas.** *Proc Natl Acad Sci U S A* 1999, **96**:1603-1608.
12. Zhang X, Gaspard JP, Chung DC: **Regulation of vascular endothelial growth factor by the Wnt and K-ras pathways in colonic neoplasia.** *Cancer Res* 2001, **61**:6050-6054.
13. 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.
14. Kendrick H, Regan JL, Magnay FA, Grigoriadis A, Mitsopoulos C, Zvelebil M, Smalley MJ: **Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate.** *BMC Genomics* 2008, **9**:591.
 15. Anastas JN, Moon RT: **WNT signalling pathways as therapeutic targets in cancer.** *Nat Rev Cancer* 2013, **13**:11-26.
 16. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
 17. 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.
 18. CancerGenomeAtlasNetwork: **Comprehensive molecular portraits of human breast tumours.** *Nature* 2012, **490**:61-70.
 19. Khramtsov AI, Khramtsova GF, Tretiakova M, Huo D, Olopade OI, Goss KH: **Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome.** *Am J Pathol* 2010, **176**:2911-2920.
 20. Segditsas S, Tomlinson I: **Colorectal cancer and genetic alterations in the Wnt pathway.** *Oncogene* 2006, **25**:7531-7537.
 21. Suzuki H, Toyota M, Carraway H, Gabrielson E, Ohmura T, Fujikane T, Nishikawa N, Sogabe Y, Nojima M, Sonoda T, et al: **Frequent epigenetic inactivation of Wnt antagonist genes in breast cancer.** *Br J Cancer* 2008, **98**:1147-1156.
 22. Veeck J, Geisler C, Noetzel E, Alkaya S, Hartmann A, Knuchel R, Dahl E: **Epigenetic inactivation of the secreted frizzled-related protein-5 (SFRP5) gene in human breast cancer is associated with unfavorable prognosis.** *Carcinogenesis* 2008, **29**:991-998.
 23. Green JL, La J, Yum KW, Desai P, Rodewald LW, Zhang X, Leblanc M, Nusse R, Lewis MT, Wahl GM: **Paracrine Wnt signaling both promotes and inhibits human breast tumor growth.** *Proc Natl Acad Sci U S A* 2013, **110**:6991-6996.
 24. Sharpless NE, Depinho RA: **The mighty mouse: genetically engineered mouse models in cancer drug development.** *Nat Rev Drug Discov* 2006, **5**:741-754.
 25. Li Y, Hively WP, Varmus HE: **Use of MMTV-Wnt-1 transgenic mice for studying the genetic basis of breast cancer.** *Oncogene* 2000, **19**:1002-1009.

26. 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.
27. Cleary AS, Leonard TL, Gestl SA, Gunther EJ: **Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers.** *Nature* 2014, **508**:113-117.
28. Zhang M, Tsimelzon A, Chang CH, Fan C, Wolff A, Perou CM, Hilsenbeck SG, Rosen JM: **Intratumoral Heterogeneity in a Trp53-Null Mouse Model of Human Breast Cancer.** *Cancer Discov* 2015.
29. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, et al: **Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts.** *Genome Biol* 2013, **14**:R125.
30. Hu Z, Fan C, Livasy C, He X, Oh DS, Ewend MG, Carey LA, Subramanian S, West R, Ikpatt F, et al: **A compact VEGF signature associated with distant metastases and poor outcomes.** *BMC Med* 2009, **7**:9.
31. Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, Perou CM: **Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures.** *BMC Med Genomics* 2011, **4**:3.
32. Collu GM, Hidalgo-Sastre A, Brennan K: **Wnt-Notch signalling crosstalk in development and disease.** *Cell Mol Life Sci* 2014, **71**:3553-3567.
33. Nishita M, Hashimoto MK, Ogata S, Laurent MN, Ueno N, Shibuya H, Cho KW: **Interaction between Wnt and TGF-beta signalling pathways during formation of Spemann's organizer.** *Nature* 2000, **403**:781-785.
34. Shackleford GM, MacArthur CA, Kwan HC, Varmus HE: **Mouse mammary tumor virus infection accelerates mammary carcinogenesis in Wnt-1 transgenic mice by insertional activation of int-2/Fgf-3 and hst/Fgf-4.** *Proc Natl Acad Sci U S A* 1993, **90**:740-744.
35. Hu T, Li C: **Convergence between Wnt-beta-catenin and EGFR signaling in cancer.** *Mol Cancer* 2010, **9**:236.
36. Pfefferle AD, Spike BT, Wahl GM, Perou CM: **Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant chemotherapy.** *Breast Cancer Res Treat* 2015, **149**:425-437.

37. Moon BS, Jeong WJ, Park J, Kim TI, Min do S, Choi KY: **Role of oncogenic K-Ras in cancer stem cell activation by aberrant Wnt/beta-catenin signaling.** *J Natl Cancer Inst* 2014, **106**:djt373.
38. AmericanCancerSociety: **Cancer Facts and Figures.** 2015.
39. Jordan VC: **Tamoxifen: a most unlikely pioneering medicine.** *Nat Rev Drug Discov* 2003, **2**:205-213.
40. Hynes NE, Lane HA: **ERBB receptors and cancer: the complexity of targeted inhibitors.** *Nat Rev Cancer* 2005, **5**:341-354.
41. Curigliano G, Goldhirsch A: **The triple-negative subtype: new ideas for the poorest prognosis breast cancer.** *J Natl Cancer Inst Monogr* 2011, **2011**:108-110.
42. Breivik J, Meling GI, Spurkland A, Rognum TO, Gaudernack G: **K-ras mutation in colorectal cancer: relations to patient age, sex and tumour location.** *Br J Cancer* 1994, **69**:367-371.
43. 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.
44. Herschkowitz JI, Zhao W, Zhang M, Usary J, Murrow G, Edwards D, Knezevic J, Greene SB, Darr D, Troester MA, et al: **Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells.** *Proc Natl Acad Sci U S A* 2012, **109**:2778-2783.
45. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L, Karginova O, Jordan J, Combest A, Bridges A, et al: **Predicting drug responsiveness in human cancers using genetically engineered mice.** *Clin Cancer Res* 2013, **19**:4889-4899.

CHAPTER 6: DISCUSSION

In 2014, over 230,000 women were diagnosed with new cases of breast cancer in the United States [1]. Given the prevalence of this disease, it is estimated that about 1 in 8 women will develop breast cancer at some point in their life [1]. Unfortunately, these pessimistic statistics routinely overshadow the optimistic fact that breast cancer related deaths have been on the decline since the early 1990s [1]! Decreasing death rates are thought to be due to a combination of factors. First, there has been an increase in public awareness due to organizations such as Susan G. Komen which spends more than 40% of their annual budget on health education. This strong advocacy effort has lead to increased screening, which has in turn resulted in more breast tumors being caught at early stages in the progression process. Second, there has been an increase in breast cancer research. Between 1930 and 1990, the death rate per 100,000 females was a little more than 30 [1]. Today, that rate is a little more than 20 per 100,000 females [1], about a 30% drop in 25 years! This is highlighted by the fact that there are about 3,000,000 breast cancer survivors living in the United States today [1].

Improved treatment of breast cancer patients is in large part due to our increased understanding that breast cancer is not a single disease, but a group of diseases [2]. Specifically, breast cancer is defined by several therapeutic subtypes ($ER^+/HER2^-$, $HER2^+$, and triple negative), and several related genomic subtypes called the intrinsic subtypes [2]. The large drop in death rate over the last 25 years [1] is due to the development of targeted

therapeutics against estrogen receptor positive [3] (luminal A/B [4]) and human epidermal growth factor receptor 2 positive [5] (HER2-enriched [4]) tumors, common subtypes of breast cancer. While this is great news, there are still about 40,000 breast cancer related deaths each year [1]. More research is therefore needed to develop therapeutics for those patients that do not respond to current standard-of-care. This is especially true for individuals with triple negative breast cancer (TNBC) (basal-like and claudin-low [4]), for whom targeted treatments are not a current option.

To develop improved therapies for TNBC, a research emphasis has been placed on determining the molecular drivers of basal-like and claudin-low tumors and to identify novel drugs that target these two subtypes. Although the genomic era promised to quickly identify these drug targets, it has also produced a large number of false positives. As a result of this and other factors, only about 5% of oncology drugs that enter clinical testing are ultimately approved by the FDA for use [6]. From a financial perspective, this low success rate means that it cost more than \$1,000,000,000 of research and development spending to develop a single oncology drug [7]. Needless to say, improved methods are needed to streamline the drug development process and to lower healthcare costs. In addition, there is a great biological need to target TNBC given its paucity of targeted therapeutic options.

Genetically engineered mouse models are a biologically relevant resource for studying mammary cancers *in vivo* under genetically controlled and immune competent conditions and may be able to bridge this gap between bench and bedside [8]. For this to happen, however, the strengths and limitations of each model need to be determined so that the models that most faithfully represent the human condition are used preclinically. This is an important first step that is unfortunately overlooked by many studies. Assuming that a

mouse model is reflective of a given human disease state is likely to lead to mistakes when interpreting experimental results. For example, it should be taking into account that the MMTV-Neu model was identified to model luminal A tumors and not HER2-Enriched tumors as would easily be assumed by the nature of the model [9].

For these reasons, we sought in Chapter 1 to highlight those murine models that most faithfully mimic the human disease state on a transcriptome level, as these models are likely to be useful for preclinical studies [9]. To ensure our analysis included as many models as possible, we consolidated 27 murine models of breast carcinoma into the largest comprehensive genomic dataset at the time of publication [9]. It should be noted that this study was only able to include so many of the most widely used breast cancer models because of collaboration and generous donations from other researchers. From our results, we were able to identify eight human-to-murine counterparts and provide insight into the molecular pathways involved in specific human breast cancer subtypes. As expected by the large degree of heterogeneity of human tumors, this study shows that multiple GEMMs are needed to represent the diversity of human breast cancers. Importantly, there was at least one murine class/model for each of the six intrinsic human subtypes analyzed [9]. These reported *trans*-species associations should guide model selection during preclinical study design to ensure appropriate representatives of human disease subtypes are used. Lastly, this study also highlights a methodology to improve preclinical study designs using mouse models for any disease, which we suggest will increase the predictive nature of preclinical studies in mice [9].

The human-to-murine subtype associations observed in Chapter 1 could be the result of two interrelated possibilities. The first possibility is that both the human and murine

subtypes share many of the same genetic drivers. Those drivers could be influencing the expression of the same set of downstream genes and as a result, an association is observed across the two species. The second possibility is that the human and murine subtypes have different genetic drivers, but both drivers independently produce an expansion of the same mammary cell type, thus possibly reflecting transformation of a common developmental stage. Although tumorigenesis will influence the expression of a subset of genes within a cell, the majority of genes that define a given cell type are likely to remain unchanged. As a result, the ‘passenger’ genes that define a given cell type are driving the observed association across the two species. This second possibility is particularly troublesome when you consider that most of the oncogene-driven mouse models analyzed use either the MMTV or WAP promoter in their design. If the activity of these promoters varies as a function of specific mammary cell types, such as luminal versus myoepithelial cells, then only those cells that naturally use these promoters would ever give rise to a tumor in these models. This is a major potential caveat that unfortunately our dataset was not powered to address.

In Chapter 2, we sought to begin addressing some of these difficult to answer questions concerning potential differences in developmental states. Specifically, we used transcriptomic profiles coming from fluorescence-activated cell sorted (FACS) normal mammary epithelial cell types from several independent human and murine studies to determine if the human-murine counterparts identified in Chapter 1 share similar normal epithelial cell features [10]. Using a meta-analysis approach, we derived consensus gene signatures for both species from normal epithelial cell types (luminal, etc) and used these to relate tumors to normal mammary epithelial cell phenotypes. Through this process, we showed that a subset of the human-murine subtype counterparts share similar normal cell

features. We proposed that molecular features associated with particular cell types along the mammary gland cell hierarchy may contribute to the clinical heterogeneity observed in human breast carcinomas. To test this, we compiled a dataset of 702 neoadjuvant anthracycline and taxane chemotherapy treated patients to determine if these cell type specific gene signatures were predictive of chemotherapy response [10]. We found that both human luminal progenitor, as well as mouse fetal mammary stem cell features, predicted pathologic complete response sensitivity across all breast cancer patients treated with neoadjuvant chemotherapy even after controlling for intrinsic subtype, proliferation, and clinical variables.

Even though targeted treatment options are improving for patients, chemotherapy still remains an important tool for clinicians. Chemotherapy is a devastating approach to fighting cancer, causing memory problems, depression, weight loss, and nausea to name a few. Being fully informed on the likely outcome of chemotherapy treatment is incredibly important for both the physician and the patient, as this information is critical when determining the next best treatment steps. Currently, physicians are forced to inform their patients that only about 15-20% of breast cancer patients will completely respond to chemotherapy treatment [11]. While this is great news for those that completely respond, the majority of patients have residual disease and are left to suffer the emotional and physical side effects caused by chemotherapy treatment.

The development of a robust laboratory test that can determine which tumors are most likely to respond to chemotherapy is greatly needed in part because it will empower patients to make more informed treatment decisions with their physicians. If it is determined that a patient has a lower chance to respond, they may not be willing to take on the side

effects of chemotherapy treatment for such a small chance of it working. This is particularly true for senior citizens who may not be as well prepared to recover from chemotherapy treatment as younger patients. Although more validation is required, our results identify several sets of good candidate genes from which to start developing this type of chemotherapy predictive test.

Even if the subset of patients that will respond to chemotherapy can be identified using our gene signatures, targeted treatment options promise to have fewer side effects. Targeted treatment options for patients with basal-like tumors remain an important unmet clinical need; therefore, we sought to identify novel genetic drivers of these tumors using a p53null mammary transplant model in Chapter 3. This model is particularly relevant because somatic p53 mutations are one of the most common genetic events in basal-like tumors [2]. Using a combination of microarray and sequencing technologies, we identified several candidate drivers of tumorigenesis in this mouse model of basal-like breast cancer. To narrow the list of candidates, a comparative analysis with human basal-like tumors was performed, and MET was found to be one of the most promising candidates. First, *MET* had a high Pearson correlation between its DNA copy number abundance and its gene expression in our mouse p53null basal-like mouse tumors only, suggesting a causal relationship. Second, *MET* is amplified in about 20% of human basal-like tumors, a conserved genetic feature across species. Third and possibly the most important, there is already a FDA approved drug that inhibits MET. Specifically, Crizotinib is approved for use in for non-small-cell lung cancer [12]. Only the *MET* amplified tumor responded to treatment, completely regressing at the end of the treatment period. This result suggests that Crizotinib may be effective as a single agent in human basal-like tumors with *MET* amplification. Since not all basal-like

patients have these aberrations, companion diagnostic tests will be required to determine which patients should receive treatment. The DNA amplification status of *MET* may serve as an excellent guide for treatment decisions, but proper clinical studies will be required to determine if that is the case. This study identifies MET as a potential driver of human basal-like breast cancer and highlights the importance of comparative genomic studies for discovering novel drug targets and for determining which patient populations are most likely to respond to treatment.

In addition to MET, Wnt signaling has also been proposed to be an attractive drug target for the basal-like subtype given the finding that canonical Wnt signaling is both enriched and predictive of poor clinical outcome in these tumors [13]. The development of effective Wnt based therapeutics, however, has been slowed in part by a limited understanding of the context dependent nature with which these aberrations influence breast tumorigenesis. In Chapter 4, we investigated our finding from Chapter 1 that MMTV-Wnt1 mice develop two classes of tumors that differ in tumor latency: Wnt1-Early^{Ex} and Wnt1-Late^{Ex} [9]. To investigate the phenotypic differences between these two subtypes we used a combination of histology, fluorescence-activated cell sorting (FACS), limiting dilution assays, mutation analysis, gene expression profiling, and drug treatments. This study definitely shows that the MMTV-Wnt1 mouse model produces two phenotypically distinct subtypes of mammary tumors. Importantly, only the Wnt1-Early^{Ex} tumors responded to Erlotinib and Lapatinib, two EGFR inhibitors. Our results suggest that this response might be related to crosstalk between canonical Wnt and EGFR signaling pathways [14]. This is important because if similar mechanisms occur in human tumors, a subset of basal-like tumors may also be sensitive to these targeted drugs.

In conclusion, the work presented here highlights several genetically engineered mouse models that are counterparts for specific human subtypes [9]. In addition to helping determine the molecular etiology of human breast tumors, these experiments also highlight several genes and pathways that may serve as attractive drug targets. These potential targets that were conserved between murine and human basal-like and claudin-low tumor counterparts are particularly interesting and important given the clinical need for targeted agents for these two intrinsic subtypes. Now that these human-to-murine counterparts have been properly identified, we hypothesize that these models should be better predictors of clinical trial success. As such, we propose that two types of preclinical studies using these mouse models should be performed simultaneously to expedite the development of improved therapeutic approaches. First, for drugs that are currently in clinical trials, murine studies should be designed to mirror these human trials. This design setup should directly inform their human trial counterparts. Second and possibly more important, novel drug regimens should also be tested using these models. This is important because while it is impractical to test an endless combination of drugs in proper clinical trials, a wide variety of drug combinations can be relatively easily tested in mice. Given that tumors are able to evade single-agent drug treatments through kinase reprogramming [15], these novel drug combinations are likely to be essential for targeting hard to treat breast tumors. If effective drug combinations against these murine models are discovered, proper clinical trials should then be performed to test the effectiveness of these treatments against human breast tumors.

REFERENCES

1. AmericanCancerSociety: **Cancer Facts and Figures**. 2015.
2. CancerGenomeAtlasNetwork: **Comprehensive molecular portraits of human breast tumours**. *Nature* 2012, **490**:61-70.
3. Jordan VC: **Tamoxifen: a most unlikely pioneering medicine**. *Nat Rev Drug Discov* 2003, **2**:205-213.
4. Prat A, Perou CM: **Deconstructing the molecular portraits of breast cancer**. *Mol Oncol* 2011, **5**:5-23.
5. Hynes NE, Lane HA: **ERBB receptors and cancer: the complexity of targeted inhibitors**. *Nat Rev Cancer* 2005, **5**:341-354.
6. Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nat Rev Drug Discov* 2004, **3**:711-715.
7. Scannell JW, Blanckley A, Boldon H, Warrington B: **Diagnosing the decline in pharmaceutical R&D efficiency**. *Nat Rev Drug Discov* 2012, **11**:191-200.
8. Sharpless NE, Depinho RA: **The mighty mouse: genetically engineered mouse models in cancer drug development**. *Nat Rev Drug Discov* 2006, **5**:741-754.
9. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, et al: **Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts**. *Genome Biol* 2013, **14**:R125.
10. Pfefferle AD, Spike BT, Wahl GM, Perou CM: **Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant chemotherapy**. *Breast Cancer Res Treat* 2015, **149**:425-437.
11. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, Ollila DW, Sartor CI, Graham ML, Perou CM: **The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes**. *Clin Cancer Res* 2007, **13**:2329-2334.
12. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, Felip E, Cappuzzo F, Paolini J, Usari T, et al: **First-line crizotinib versus chemotherapy in ALK-positive lung cancer**. *N Engl J Med* 2014, **371**:2167-2177.
13. Khramtsov AI, Khramtsova GF, Tretiakova M, Huo D, Olopade OI, Goss KH: **Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome**. *Am J Pathol* 2010, **176**:2911-2920.

14. Hu T, Li C: **Convergence between Wnt-beta-catenin and EGFR signaling in cancer.** *Mol Cancer* 2010, **9**:236.
15. Duncan JS, Whittle MC, Nakamura K, Abell AN, Midland AA, Zawistowski JS, Johnson NL, Granger DA, Jordan NV, Darr DB, et al: **Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer.** *Cell* 2012, **149**:307-321.

APPENDIX A: RELATED COAUTHORSHIPS

Cancer Res. 2012 Jul 1;72(13):3260-9.

Oncogenic PI3K mutations lead to NF- κ B-dependent cytokine expression following growth factor deprivation

Hutti JE, Pfefferle AD, Russell SC, Sircar M, Perou CM, Baldwin AS.

Abstract: The phosphoinositide 3-kinase (PI3K) pathway is one of the most commonly misregulated signaling pathways in human cancers, but its impact on the tumor microenvironment has not been considered as deeply as its autonomous impact on tumor cells. In this study, we show that NF- κ B is activated by the two most common PI3K mutations, PIK3CA E545K and H1047R. We found that markers of NF- κ B are most strongly upregulated under conditions of growth factor deprivation. Gene expression analysis conducted on cells deprived of growth factors identified the repertoire of genes altered by oncogenic PI3K mutations following growth factor deprivation. This gene set most closely correlated with gene signatures from claudin-low and basal-like breast tumors, subtypes frequently exhibiting constitutive PI3K/Akt activity. An NF- κ B-dependent subset of genes driven by oncogenic PI3K mutations was also identified that encoded primarily secreted proteins, suggesting a paracrine role for this gene set. Interestingly, while NF- κ B activated by oncogenes such as Ras and EGF receptor leads to cell-autonomous effects, abrogating NF- κ B in PI3K-transformed cells did not decrease proliferation or induce apoptosis. However, conditioned media from PI3K mutant-expressing cells led to increased STAT3 activation in recipient THP-1 monocytes or normal epithelial cells in a NF- κ B and interleukin-6-dependent manner. Together, our findings describe a PI3K-driven, NF- κ B-dependent transcriptional profile that may play a critical role in promoting a microenvironment amenable to tumor progression. These data also indicate that NF- κ B plays diverse roles downstream from different oncogenic signaling pathways.

LKB1/STK11 inactivation leads to expansion of a pro-metastatic tumor sub-population in melanoma

Liu W, Monahan KB, **Pfefferle AD**, Shimamura T, Sorrentino J, Chan K, Roadcap DW, Ollila DW, Nancy NE, Castrillon DH, Miller CR, Perou CM, Wong KK, Bear JE, Sharpless NE.

Abstract: Germline mutations in LKB1 (STK11) are associated with the Peutz-Jeghers syndrome (PJS), which includes aberrant mucocutaneous pigmentation, and somatic LKB1 mutations occur in 10% of cutaneous melanoma. By somatically inactivating Lkb1 with K-Ras activation (\pm p53 loss) in murine melanocytes, we observed variably pigmented and highly metastatic melanoma with 100% penetrance. LKB1 deficiency resulted in increased phosphorylation of the SRC family kinase (SFK) YES, increased expression of WNT target genes, and expansion of a CD24(+) cell population, which showed increased metastatic behavior in vitro and in vivo relative to isogenic CD24(-) cells. These results suggest that LKB1 inactivation in the context of RAS activation facilitates metastasis by inducing an SFK-dependent expansion of a prometastatic, CD24(+) tumor subpopulation.

Comparative oncogenomics implicates the Neurofibromin 1 gene (*NF1*) as a breast cancer driver

Wallace MD, **Pfefferle AD**^{*}, Shen L^{*}, McNairn AJ, Cerami EG, Fallon BL, Rinaldi VD,
Southard TL, Perou CM, Schimenti JC.

Abstract: Identifying genomic alterations driving breast cancer is complicated by tumor diversity and genetic heterogeneity. Relevant mouse models are powerful for untangling this problem because such heterogeneity can be controlled. Inbred Chaos3 mice exhibit high levels of genomic instability leading to mammary tumors that have tumor gene expression profiles closely resembling mature human mammary luminal cell signatures. We genomically characterized mammary adenocarcinomas from these mice to identify cancer-causing genomic events that overlap common alterations in human breast cancer. Chaos3 tumors underwent recurrent copy number alterations (CNAs), particularly deletion of the RAS inhibitor Neurofibromin 1 (Nf1) in nearly all cases. These overlap with human CNAs including NF1, which is deleted or mutated in 27.7% of all breast carcinomas. Chaos3 mammary tumor cells exhibit RAS hyperactivation and increased sensitivity to RAS pathway inhibitors. These results indicate that spontaneous NF1 loss can drive breast cancer. This should be informative for treatment of the significant fraction of patients whose tumors bear NF1 mutations.

Combined PI3K/mTOR and MEK inhibition provides broad anti-tumor activity in faithful murine cancer models

Roberts PJ, Usary JE, Darr DB, Dillon PM, **Pfefferle AD**, Johnson SM, Combest AJ, Jin J, Zamboni WC, Perou CM, Sharpless NE.

Abstract: Anticancer drug development is inefficient, but genetically engineered murine models (GEMM) and orthotopic, syngeneic transplants (OST) of cancer may offer advantages to in vitro and xenograft systems. We assessed the activity of 16 treatment regimens in a RAS-driven, Ink4a/Arf-deficient melanoma GEMM. In addition, we tested a subset of treatment regimens in three breast cancer models representing distinct breast cancer subtypes: claudin-low (T11 OST), basal-like (C3-TAg GEMM), and luminal B (MMTV-Neu GEMM). Like human RAS-mutant melanoma, the melanoma GEMM was refractory to chemotherapy and single-agent small molecule therapies. Combined treatment with AZD6244 [mitogen-activated protein-extracellular signal-regulated kinase kinase (MEK) inhibitor] and BEZ235 [dual phosphoinositide-3 kinase (PI3K)/mammalian target of rapamycin (mTOR) inhibitor] was the only treatment regimen to exhibit significant antitumor activity, showed by marked tumor regression and improved survival. Given the surprising activity of the "AZD/BEZ" combination in the melanoma GEMM, we next tested this regimen in the "claudin-low" breast cancer model that shares gene expression features with melanoma. The AZD/BEZ regimen also exhibited significant activity in this model, leading us to testing in even more diverse GEMMs of basal-like and luminal breast cancer. The AZD/BEZ combination was highly active in these distinct breast cancer models, showing equal or greater efficacy compared with any other regimen tested in studies of over 700 tumor-bearing mice. This regimen even exhibited activity in lapatinib-resistant HER2(+) tumors. These results show the use of credentialed murine models for large-scale efficacy testing of diverse anticancer regimens and predict that combinations of PI3K/mTOR and MEK inhibitors will show antitumor activity in a wide range of human malignancies.

Conditional loss of ErbB3 delays mammary gland hyperplasia induced by mutant PIK3CA without affecting mammary tumor latency, gene expression or signaling

Young CD, **Pfefferle AD**, Owens P, Kuba MG, Rexer BN, Balko JM, Sanchez V, Cheng H, Perou CM, Zhao JJ, Cook RS, Arteaga CL.

Abstract: Mutations in PIK3CA, the gene encoding the p110 α catalytic subunit of phosphoinositide 3-kinase (PI3K), have been shown to transform mammary epithelial cells (MEC). Studies suggest this transforming activity requires binding of mutant p110 α via p85 to phosphorylated YXXM motifs in activated receptor tyrosine kinases (RTK) or adaptors. Using transgenic mice, we examined if ErbB3, a potent activator of PI3K, is required for mutant PIK3CA-mediated transformation of MECs. Conditional loss of ErbB3 in mammary epithelium resulted in a delay of PIK3CA(H1047R)-dependent mammary gland hyperplasia, but tumor latency, gene expression, and PI3K signaling were unaffected. In ErbB3-deficient tumors, mutant PI3K remained associated with several tyrosyl phosphoproteins, potentially explaining the dispensability of ErbB3 for tumorigenicity and PI3K activity. Similarly, inhibition of ErbB RTKs with lapatinib did not affect PI3K signaling in PIK3CA(H1047R)-expressing tumors. However, the p110 α -specific inhibitor BYL719 in combination with lapatinib impaired mammary tumor growth and PI3K signaling more potently than BYL719 alone. Furthermore, coinhibition of p110 α and ErbB3 potently suppressed proliferation and PI3K signaling in human breast cancer cells harboring PIK3CA(H1047R). These data suggest that PIK3CA(H1047R)-driven tumor growth and PI3K signaling can occur independently of ErbB RTKs. However, simultaneous blockade of p110 α and ErbB RTKs results in superior inhibition of PI3K and mammary tumor growth, suggesting a rational therapeutic combination against breast cancers harboring PIK3CA activating mutations.

Mutant PIK3CA accelerates HER2-driven transgenic mammary tumors and induces resistance to combinations of anti-HER2 therapies

Hanker AB, Pfefferle AD, Balko JM, Kuba MG, Young CD, Sanchez V, Sutton CR, Cheng H, Perou CM, Zhao JJ, Cook RS, Arteaga CL.

Abstract: Human epidermal growth factor receptor 2 (HER2; ERBB2) amplification and phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (PIK3CA) mutations often co-occur in breast cancer. Aberrant activation of the phosphatidylinositol 3-kinase (PI3K) pathway has been shown to correlate with a diminished response to HER2-directed therapies. We generated a mouse model of HER2-overexpressing (HER2(+)), PIK3CA(H1047R)-mutant breast cancer. Mice expressing both human HER2 and mutant PIK3CA in the mammary epithelium developed tumors with shorter latencies compared with mice expressing either oncogene alone. HER2 and mutant PIK3CA also cooperated to promote lung metastases. By microarray analysis, HER2-driven tumors clustered with luminal breast cancers, whereas mutant PIK3CA tumors were associated with claudin-low breast cancers. PIK3CA and HER2(+)/PIK3CA tumors expressed elevated transcripts encoding markers of epithelial-to-mesenchymal transition and stem cells. Cells from HER2(+)/PIK3CA tumors more efficiently formed mammospheres and lung metastases. Finally, HER2(+)/PIK3CA tumors were resistant to trastuzumab alone and in combination with lapatinib or pertuzumab. Both drug resistance and enhanced mammosphere formation were reversed by treatment with a PI3K inhibitor. In sum, PIK3CA(H1047R) accelerates HER2-mediated breast epithelial transformation and metastatic progression, alters the intrinsic phenotype of HER2-overexpressing cancers, and generates resistance to approved combinations of anti-HER2 therapies.

Endothelial-like properties of claudin-low breast cancer cells promote tumor vascular permeability and metastasis

Harrell JC, **Pfefferle AD**, Zalles N, Prat A, Fan C, Khramtsov A, Olopade OI, Troester MA, Dudley AC, Perou CM.

Abstract: The vasculature serves as the main conduit for breast tumor metastases and is a target of therapeutics in many tumor types. In this study, we aimed to determine if tumor-associated vascular properties could help to explain the differences observed in metastagenicity across the intrinsic subtypes of human breast tumors. Analysis of gene expression signatures from more than 3,000 human breast tumors found that genomic programs that measured vascular quantity, vascular proliferation, and a VEGF/Hypoxia-signature were the most highly expressed in claudin-low and basal-like tumors. The majority of the vascular gene signatures added metastasis-predictive information to immunohistochemistry-defined microvessel density scores and genomically defined-intrinsic subtype classification. Interestingly, pure claudin-low cell lines, and subsets of claudin-low-like cells within established basal-like cancer cell lines, exhibited endothelial/tube-like morphology when cultured on Matrigel. In vivo xenografts found that claudin-low tumors, but not luminal tumors, extensively perfused injected contrast agent through paracellular spaces and non-vascular tumor-lined channels. Taken together, the endothelial-like characteristics of the cancer cells, combined with both the amount and the physiologic state of the vasculature contribute to breast cancer metastatic progression. We hypothesize that the genetic signatures we have identified highlight patients that should respond most favorably to anti-vascular agents.

c-Myc and Her2 cooperate to drive stem-like phenotype with poor prognosis in breast cancer

Nair R, Roden DL, Teo WS, McFarland A, Junankar S, Ye S, Nguyen A, Yang J, Nikolic I, Hui M, Morey A, Shah J, **Pfefferle AD**, Usary J, Selinger C, Baker LA, Armstrong N, Cowley MJ, Naylor MJ, Ormandy CJ, Lakhani SR, Herschkowitz JI, Perou CM, Kaplan W, O'Toole SA, Swarbrick A.

Abstract: The HER2 (ERBB2) and MYC genes are commonly amplified in breast cancer, yet little is known about their molecular and clinical interaction. Using a novel chimeric mammary transgenic approach and in vitro models, we demonstrate markedly increased self-renewal and tumour-propagating capability of cells transformed with Her2 and c-Myc. Coexpression of both oncoproteins in cultured cells led to the activation of a c-Myc transcriptional signature and acquisition of a self-renewing phenotype independent of an epithelial-mesenchymal transition programme or regulation of conventional cancer stem cell markers. Instead, Her2 and c-Myc cooperated to induce the expression of lipoprotein lipase, which was required for proliferation and self-renewal in vitro. HER2 and MYC were frequently coamplified in breast cancer, associated with aggressive clinical behaviour and poor outcome. Lastly, we show that in HER2(+) breast cancer patients receiving adjuvant chemotherapy (but not targeted anti-Her2 therapy), MYC amplification is associated with a poor outcome. These findings demonstrate the importance of molecular and cellular context in oncogenic transformation and acquisition of a malignant stem-like phenotype and have diagnostic and therapeutic consequences for the clinical management of HER2(+) breast cancer.

Expression of miR-200c in claudin-low breast cancer alters stem cell functionality, enhances chemosensitivity and reduces metastatic potential

Knezevic J, **Pfefferle AD**, Petrovic I, Greene S, Perou CM, Rosen JM.

Abstract: Claudin-low tumors are a highly aggressive breast cancer subtype with no targeted treatments and a clinically documented resistance to chemotherapy. They are significantly enriched in cancer stem cells (CSCs), which makes claudin-low tumor models particularly attractive for studying CSC behavior and developing novel approaches to minimize CSC therapy resistance. One proposed mechanism by which CSCs arise is via an epithelial-mesenchymal transition (EMT), and reversal of this process may provide a potential therapeutic approach for increasing tumor chemosensitivity. Therefore, we investigated the role of known EMT regulators, miR-200 family of microRNAs in controlling the epithelial state, stem-like properties and therapeutic response in an in vivo primary, syngeneic p53^{null} claudin-low tumor model that is normally deficient in miR-200 expression. Using an inducible lentiviral approach, we expressed the miR-200c cluster in this model and found that it changed the epithelial state, and consequently, impeded CSC behavior in these mesenchymal tumors. Moreover, these state changes were accompanied by a decrease in proliferation and an increase in the differentiation status. miR-200c expression also forced a significant reorganization of tumor architecture, affecting important cellular processes involved in cell-cell contact, cell adhesion and motility. Accordingly, induced miR200c expression significantly enhanced the chemosensitivity and decreased the metastatic potential of this p53^{null} claudin-low tumor model. Collectively, our data suggest that miR-200c expression in claudin-low tumors offers a potential therapeutic application to disrupt the EMT program on multiple fronts in this mesenchymal tumor subtype, by altering tumor growth, chemosensitivity and metastatic potential in vivo.

JNK2 prevents luminal cell commitment in normal mammary glands and tumors by inhibiting p53/NOTCH1 and BRCA1 expression

Cantrell MA, Ebelt ND, **Pfefferle AD**, Perou CM, Van Den Berg CL.

Abstract: Breast cancer is a heterogeneous disease with several subtypes carrying unique prognoses. Patients with differentiated luminal tumors experience better outcomes, while effective treatments are unavailable for poorly differentiated tumors, including the basal-like subtype. Mechanisms governing mammary tumor subtype generation could prove critical to developing better treatments. C-Jun N-terminal kinase 2 (JNK2) is important in mammary tumorigenesis and tumor progression. Using a variety of mouse models, human breast cancer cell lines and tumor expression data, studies herein support that JNK2 inhibits cell differentiation in normal and cancer-derived mammary cells. JNK2 prevents precocious pubertal mammary development and inhibits Notch-dependent expansion of luminal cell populations. Likewise, JNK2 suppresses luminal populations in a p53-competent Polyoma Middle T-antigen tumor model where *jnk2* knockout causes p53-dependent upregulation of *Notch1* transcription. In a *p53* knockout model, JNK2 restricts luminal populations independently of Notch1, by suppressing *Brcal* expression and promoting epithelial to mesenchymal transition. JNK2 also inhibits estrogen receptor (ER) expression and confers resistance to fulvestrant, an ER inhibitor, while stimulating tumor progression. These data suggest that therapies inhibiting JNK2 in breast cancer may promote tumor differentiation, improve endocrine therapy response, and inhibit metastasis.

APPENDIX B: CURRICULUM VITAE

ADAM D. PFEFFERLE

97 Grapevine Trail, Durham NC 27707 · 704.564.4921 · adamp@email.unc.edu
<http://www.linkedin.com/in/adampfefferle>

EDUCATION

Ph.D.	Molecular and Cellular Pathology University of North Carolina at Chapel Hill, Chapel Hill NC Advisor: Charles M. Perou, Ph.D.	2015
M.Eng.	Engineering North Carolina State University, Raleigh NC Concentration: Materials Science and Engineering	2014
EMT-I	Emergency Medical Technician Alamance Community College, Haw River NC Certifications: Basic and Intermediate	2009
B.S.	Chemistry and Biology University of North Carolina at Chapel Hill, Chapel Hill NC Advisor: Susan T. Lord, Ph.D.	2006

RESEARCH AND WORK EXPERIENCE

University of North Carolina at Chapel Hill: Graduate Student Department of Pathology and Laboratory Medicine Focus: Intrinsic Subtypes of Breast Carcinoma Advisor: Charles M. Perou, Ph.D.	2009-2015
Duke University: Research Technician Institute for Genome Science & Policy Focus: Evolution of Transcriptional Regulation Advisor: Gregory A. Wray, Ph.D.	2008-2009
Cogenics: Associate Scientist II Perimeter Park, Morrisville NC Focus: Microarray Processing Microarray Division	2006-2008

PUBLICATIONS

1. Cantrell MA, Ebelt ND, **Pfefferle AD**, Perou CM, Van Den Berg CL. JNK2 prevents luminal cell commitment in normal mammary glands and tumors by inhibiting p53/NOTCH1 and BRCA1 expression. *Oncotarget*. **2015**.
2. **Pfefferle AD**, Spike BT, Wahl GM, Perou CM. Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant anthracycline and taxane chemotherapy. *Breast Cancer Research and Treatment*. **2015**.
3. Knezevic J, **Pfefferle AD**, Petrovic I, Greene S, Perou CM, Rosen JM. Expression of miR-200c in claudin-low breast cancer alters stem cell functionality, enhances chemosensitivity and reduces metastatic potential. *Oncogene*. **2015**.
4. Nair R, Roden DL, Teo WS, McFarland A, Junankar S, Ye S, Nguyen A, Yang J, Nikolic I, Hui M, Morey A, Shah J, **Pfefferle AD**, Usary J, Selinger C, Baker LA, Armstrong N, Cowley MJ, Naylor MJ, Ormandy CJ, Lakhani SR, Herschkowitz JI, Perou CM, Kaplan W, O'Toole SA, Swarbrick A. c-Myc and Her2 cooperate to drive stem-like phenotype with poor prognosis in breast cancer. *Oncogene*. **2014**.
5. Harrell JC, **Pfefferle AD**, Zalles N, Prat A, Fan C, Khramtsov A, Olopade OI, Troester MA, Dudley AC, Perou CM. Endothelial-like properties of claudin-low breast cancer cells promote tumor vascular permeability and metastasis. *Clinical and Experimental Metastasis*. **2014**.
6. **Pfefferle AD**, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, Rosen JM, Perou CM. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biology*. **2013**.
7. Hanks AB, **Pfefferle AD**, Balko JM, Kuba MG, Young CD, Sanchez V, Sutton CR, Cheng H, Perou CM, Zhao JJ, Cook RS, Arteaga CL. Mutant PIK3CA accelerates HER2-driven transgenic mammary tumors and induces resistance to combinations of anti-HER2 therapies. *PNAS*. **2013**.

8. Young CD, **Pfefferle AD**, Owens P, Kuba MG, Rexer BN, Balko JM, Sanchez V, Cheng H, Perou CM, Zhao JJ, Cook RS, Arteaga CL. Conditional loss of ErbB3 delays mammary gland hyperplasia induced by mutant PIK3CA without affecting mammary tumor latency, gene expression or signaling. *Cancer Research*. **2013**.
9. Wallace MD, **Pfefferle AD**^{*}, Shen L^{*}, McNairn AJ, Cerami EG, Fallon BL, Rinaldi VD, Southard TL, Perou CM, Schimenti JC. Comparative oncogenomics implicates the Neurofibromin 1 gene (*NF1*) as a breast cancer driver. *Genetics*. **2012**.
10. Roberts PJ, Usary JE, Darr DB, Dillon PM, **Pfefferle AD**, Johnson SM, Combest AJ, Jin J, Zamboni WC, Perou CM, Sharpless NE. Combined PI3K/mTOR and MEK inhibition provides broad anti-tumor activity in faithful murine cancer models. *Clinical Cancer Research*. **2012**.
11. Hutti JE, **Pfefferle AD**, Russell SC, Sircar M, Perou CM, Baldwin AS. Oncogenic PI3K mutations lead to NF- κ B-dependent cytokine expression following growth factor deprivation. *Cancer Research*. **2012**.
12. Liu W, Monahan KB, **Pfefferle AD**, Shimamura T, Sorrentino J, Chan K, Roadcap DW, Ollila DW, Nancy NE, Castrillon DH, Miller CR, Perou CM, Wong KK, Bear JE, Sharpless NE. LKB1/STK11 inactivation leads to expansion of a pro-metastatic tumor sub-population in melanoma. *Cancer Cell*. **2012**.
13. Fedrigo O, **Pfefferle AD**, Babbitt CC, Haygood R, Wall CE, Wray GA. A potential role for glucose transporters in the evolution of human brain size. *Brain, Behavior, and Evolution*. **2011**.
14. **Pfefferle AD**^{*}, Warner LR^{*}, Wang CW, Nielsen WJ, Babbitt CC, Fedrigo O, Wray GA. Comparative expression analysis of the phosphocreatine circuit in extant primates: Implications for human brain evolution. *Journal of Human Evolution*. **2011**.
15. Fedrigo O, Warner LR, **Pfefferle AD**, Babbitt CC, Cruz-Gordillo P, Wray GA. A pipeline to determine RT-QPCR control genes for evolutionary studies: Application to primate gene expression across multiple tissues. *PLoS One*. **2010**.
16. Babbitt CC, Fedrigo O, **Pfefferle AD**, Boyle AP, Horvath JE, Furey TS, Wray GA. Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biology and Evolution*. **2010**.

PRESENTATIONS

Department of Pathology Grand Rounds: Invited Talk	2015
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Identifying personalized therapeutic targets	
Department of Pathology Annual Research Symposium: Poster	2014
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Predicting breast tumor response to neoadjuvant chemotherapy	
Genetics Department Seminar Series: Invited Talk	2014
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Genetically engineered mouse models of breast carcinoma	
Molecular and Cellular Pathology Seminar: Invited Talk	2014
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Genetically engineered mouse models of breast carcinoma	
Department of Pathology Annual Research Symposium: Poster	2013
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Intrinsic subtyping of genetically engineered mouse models	
Carolina Biosciences Alumni Reunion: Poster	2013
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Intrinsic subtyping of genetically engineered mouse models	
Department of Pathology Annual Research Symposium: Invited Talk	2012
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Intrinsic subtyping of genetically engineered mouse models	
Department of Pathology Annual Research Symposium: Poster	2011
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Intrinsic subtyping of genetically engineered mouse models	
Molecular and Cellular Pathology Seminar: Invited Talk	2011
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Molecular mechanisms of angiogenesis and metastasis	

Department of Pathology Annual Research Symposium: Poster	2010
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Targeted inhibition of the PI3K/mTOR pathway in breast tumors	
Biological and Biomedical Sciences Program (BBSP): Invited Talk	2010
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Using FAIRE to analyze open chromatin structure	
Winter 2009 BBSP Poster Session: Poster	2009
University of North Carolina at Chapel Hill, Chapel Hill NC	
Topic: Transcriptomic analysis of genetically engineered astrocytomas	

TEACHING AND MENTORING

Undergraduate Research Mentor	2012-2015
University of North Carolina at Chapel Hill, Chapel Hill NC	
Student: Yash N. Agrawal	
Undergraduate Research Mentor	2011-2013
University of North Carolina at Chapel Hill, Chapel Hill NC	
Student: Nicole Zalles	
Graduate Research Mentor	2010-2012
University of North Carolina at Chapel Hill, Chapel Hill NC	
Students: BBSP First Year Students	
Teaching Assistant	2010
University of North Carolina at Chapel Hill, Chapel Hill NC	
Course: Microscopic Anatomy	
Undergraduate Research Mentor	2008-2009
Duke University, Durham NC	
Student: Peter Cruz-Gordillo	
Undergraduate Research Mentor	2008-2009
Duke University, Durham NC	
Student: Catrina W. Wang	

CERTIFICATIONS

EMT-I	NC Emergency Medical Technician Intermediate	2009-2014
NIMS	National Incident Management System	2009-2014
CPR	American Heart Association	2004-2014
EMT-B	NC Emergency Medical Technician Basic	2004-2009
RBL	Regional Biocontainment Lab	2008-2009

OUTREACH

Full Member of Sigma Xi	2015-Present
South Orange Rescue Squad: EMS Division, volunteered 750+ hours	2010-2012
Health Science Camp	2012
BBSP Poster Session Judge	2010, 2011
TIBBS Student Selection Hiring Committee	2011
Home School Scientific Workshop	2009-2010
North Carolina DNA Day	2010