USING GENE EXPRESSION ANALYSIS TO GUIDE AND IDENTIFY TREATMENTS FOR BREAST CANCER PATIENTS

USING GENE EXPRESSION ANALYSIS TO GUIDE AND IDENTIFY TREATMENTS FOR BREAST CANCER PATIENTS

By ROBIN M. HALLETT, B.Sc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2012) McMaster University

Hamilton, Ontario (Biochemistry)

TITLE: USING GENE EXPRESSION ANALYSIS TO GUIDE AND IDENTIFY TREATMENTS FOR BREAST CANCER PATIENTS

AUTHOR: Robin M. Hallett, B.Sc.

SUPERVISOR: Dr. John A. Hassell

NUMBER OF PAGES: XVI, 220

ABSTRACT

Based on breast cancer clinical trial data accumulated over the last several decades it is obvious that standard breast cancer therapeutics, such as tamoxifen and cytotoxic chemotherapy, extend survival in breast cancer patients. However, it has also become obvious that only a minority of patients within these trials derive benefit from treatment. In a population of breast cancer patients treated with adjuvant therapy after surgery, many patients are over-treated, as they would never experience relapse even without receiving adjuvant therapies. Among the remaining patients, some achieve durable remission from therapy, whereas others relapse despite therapy. Hence, there is an obvious need to develop biomarkers that can serve to identify these three populations of patients, such that only patients who are likely to benefit from available therapies are treated with these therapies, as well as to develop new therapies for the treatment of patients who aren't afforded durable remission by approved treatments. In this thesis we present the identification of biomarkers that can be used to identify low risk breast cancer patients who experience excellent long-term survival even without adjuvant therapy. Conversely, patients identified as high risk represent those patients most likely to benefit from intervention with aggressive treatment regimens. We also report on the identification of biomarkers which can predict the likelihood of response to approved cytotoxic chemotherapy regimens, which could be used to further refine stratification of high risk patients into those whose tumors would likely respond to approved therapies. Finally, for high risk patients unlikely to be afforded durable remission from available

therapies, we report on the identification of agents that target breast tumor-initiating cells, and may be effective for the treatment of these patients.

ACKNOWLEGEMENTS

First, I would like to acknowledge and thank my Ph.D. supervisor Dr. John Hassell for his mentorship during, and contribution to my 'scientific maturation'. It's hard to imagine an environment that would have been more suitable for scientific development during my studies. This was an environment that allowed me to fail, but also to flourish (hopefully more than fail) and where I was allowed to work on numerous projects that encompassed many aspects of breast cancer research. I also had a lot of fun. I would also like to acknowledge and thank Dr. Brad Doble, a member of my supervisory committee for helpful feedback and making time to talk with me whenever required. I also thank Dr Anita Bane, the other member of my supervisory committee and collaborator for many helpful discussions, and for providing doorway to pursue translation research projects with actual breast cancer patient samples.

I would also like to acknowledge and thank Drs. Johnathan Bramson and Sheila Singh for letting me work with them on non-breast cancer projects, which was very interesting.

I would like to acknowledge and thank all past and present members of the Hassell lab, especially Maria, Andrew, Allison, Craig, Adele and Bonnie, who all made different but important contributions to the scientist I've become.

Finally, I would like to thank nature, for making such an interesting and complicated world. I don't know what I'd spend my time doing if the world wasn't so hard to understand.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Manuscript #1

Manuscript #2

Manuscript #3

Manuscript #4

Manuscript #5

PREFACE AND DECLARATION OF ACADEMIC ACHIEVEMENT

The following thesis comprises 5 manuscripts prepared as a sandwich thesis. The first (Sci. Rep., 2011), third (BMC Medical Genomics) and fifth (PLoS ONE, 2012) manuscripts are published whereas the second (PLoS Gen.) and fourth (Nat. Chem. Biol.) manuscripts are at various pre-publication stages. The introduction from each manuscript has been removed to avoid overlapping content with the introductory chapter. However, each chapter is preceded by a short introduction to discuss the pertinent implications of the work as well as provide a very brief but up-to-date synopsis of the field. The manuscript chapter references are self-contained within each of the manuscript chapters, and references for the remainder of the thesis are included at the end of the final chapter. The final chapter is meant to draw attention to the challenges facing breast cancer patient treatment and how the contents of this thesis attempt to address these challenges.

LIST OF ABBREVIATIONS

ADF: Actin Depolymerisation Factor

ALL: Acute Lymphoblastic Leukemia

AML: Acute Myeloid Leukemia

ANOVA: Analysis of Variance

AUC: Area Under Curve

BLBC: Basal-like Breast Cancer

BTIC: Breast Tumor-Initiating Cell

CCAC: Cancer Centre for Animal Control

CK: Cytokeratin

DFS: Disease Free Survival

DMFS: Distant Metastasis Free Survival

DMSO: Dimethyl Sulfoxide

DNA: Deoxyribonucleic Acid

EGF: Epidermal Growth Factor

EGFR: Epidermal Growth Factor Receptor

EMT: Epithelial Mesenchymal Transition

ER: Estrogen Receptor

FACS: Fluorescent Activated Cell Sorting

FBS: Fetal Bovine Serum

FDA: Food and Drug Administration

FGF: Fiboblast Growth Factor

fRMA: Frozen Robust Multichip Algorithm

FVB: Friend Virus B Tropism

GEO: Gene Expression Omnibus

H&E: Hematoxylin and Eosin

HER2: Human Epidermal Growth Factor Receptor 2

HER3: Human Epidermal Growth Factor Receptor 3

HR: Hazard Ratio

IC50: Inhibitor Concentration 50

IGFR: Insulin-like Growth Factor 1 Receptor

LOWESS: Locally Weighted Scatter Plot Smoothing

mAChR: Muscarinic Acetylcholine Receptor

MINDACT: Microarray In Node Negative and 1-3 positive lymph node Disease may Avoid Chemotherapy

MMTV: Mouse Mammary Tumor Virus

nAChR: Nicotinic Acetylcholine Receptor

NCI: National Cancer Institute

NMF: Negative Matrix Factorization

NSABP: National Surgical Adjuvant Breast and Bowel Project

PAM: Prediction Analysis Microarrays

PARP: Poly-ADP Ribose Polymerase

pCR: Pathological Complete Respsonse

PR: Progesterone Receptor

PTEN: Phosphatase and Tensin Homolog

RCB: Residual Cancer Burden

RD: Residual Disease

RMA: Robust Multichip Algorithm

RNA: Ribonucleic Acid

ROC: Receiver Operator Characteristic Curve

RS: Recurrence Score

RT-PCR: Reverse Transcriptase – Polymerase Chain Reaction

TFAC: Paclitaxel/Docetaxel, 5-fluorouracil, Adriamycin and Cyclophosphamide

TIC: Tumor-initiating Cell

CHAPTER 1

Introduction

Breast cancer

Breast cancer continues to be a leading cause of cancer mortality in Canada (Canadian Cancer Society, 2011). Whereas our understanding of breast cancer continues to increase, these advances have rarely translated into improved treatments for breast cancer patients. With the exception of trastuzumab (Herceptin^{TM}) to treat patients with HER2 or ERBB2 overexpressing tumors (Slamon *et al.*, 1989; Vogel *et al.*, 2002), endocrine therapy and cytotoxic chemotherapy remain the mainstays for breast cancer treatment. Notably, in many cases these therapies do not provide any clinical benefit to patients (Fisher *et al.*, 1989; Fisher *et al.*, 1997; Fisher *et al.*, 2004). Moreover, administration of chemotherapy, particularly cytotoxic agents, results in patients experiencing adverse side-effects and a significant decrease in their quality of life (Herbst *et al.*, 2006). Clearly there is a significant need for breast cancer research, where the findings of the research may be rapidly translated into improved clinical outcomes and higher quality of life for breast cancer patients.

Gene expression signatures in breast cancer

The advent of gene expression profiling technologies enables high-throughput analysis of transcript abundance for thousands of transcripts in parallel. Such, studies have identified 'intrinsic' molecular subtypes of breast cancer, confirming that breast cancer is not a single disease, but rather represents a collection of neoplastic disorders

that share a common origin in breast epithelium (Gatza *et al.*, 2010; Parker *et al.*, 2009; Perou *et al.*, 2000; Sorlie *et al.*, 2001; Sorlie *et al.*, 2003; Sotiriou and Pusztai, 2009). It is possible to distill the gene expression profiles of human breast tumors or breast tumor cell lines, into relatively short gene lists, commonly known as gene signatures, which can serve as biomarkers for particular biological processes or clinical attributes. Gene signatures have already been identified with the capacity to predict the outcome of breast cancer patients and to identify the activity of various oncogenic signaling pathways, as well as to indicate response to commonly used anticancer drugs (Bild *et al.*, 2006; Bogaerts *et al.*, 2006; Buyse *et al.*, 2006; Chang *et al.*, 2005; Dai *et al.*, 2005; Desmedt *et al.*, 2008; Hallett *et al.*, 2010; Hassane *et al.*; Ivshina *et al.*, 2006; Lamb *et al.*, 2006; Reis-Filho *et al.*, 2006; Staunton *et al.*, 2001; Vuaroqueaux *et al.*, 2007; Yau *et al.*, 2010). In short, gene expression signatures hold great promise to increase our understanding of breast cancer biology and accelerate the development of tailored therapy regiments based on the molecular characteristics of an individual's breast tumor.

Ultimately, gene expression data can be harnessed to provide more effective management and treatment of breast cancer patients on many levels. For example, prognostic signatures could be used as biomarkers to identify low-risk patients and spare them harmful and unnecessary chemotherapy, as well as to identify higher risk patients who would require more aggressive therapy. Within high risk patients, predictive signatures could be used as biomarkers to learn what regimen might be most beneficial to that particular patient. Moreover, for patients identified as being unlikely to benefit from

available therapies, and for whom novel therapies are required, gene expression analysis can aid to identifying the next generation of experimental therapies.

Molecular classification of breast cancer

Multiple molecular subtypes of breast cancer were originally described in 2000 from gene expression profiling studies completed on invasive breast cancers (Perou *et al.*, 2000). In general, 4 molecular subtypes have been reproducibly identified, including luminal A tumors, which are mostly estrogen receptor (ER)-positive and of low grade, a pathological measurement of differentiation and proliferation; luminal B tumors, which are also mostly ER-positive but often of high grade; basal-like breast tumors which roughly correspond to ER-negative, progesterone receptor (PR)- negative, and HER2-negative tumors (also commonly called 'triple negative'); and HER2+ breast tumors that have amplification and/or overexpression of the *HER2* gene (Hu *et al.*, 2006; Perou *et al.*, 2000; Sorlie, 2004; Sorlie *et al.*, 2001; Sorlie *et al.*, 2003; Sotiriou *et al.*, 2003). A normal breast-like subgroup has also been identified that expresses genes characteristic of adipose tissue suggesting that this subgroup may be a technical artifact resulting from low tumor cellularity. Importantly, these subtypes roughly coincide with standard histopathological classification on the basis or ER, PR, and HER2, as well as grade.

Basal-like breast cancers do not generally express the ER, PR or HER2, but express the cytokeratins (CK) 5, and/or CK 17, which are characteristic of the basal/myoepithelial cell layer of the normal breast epithelium, as well as elevated levels of growth factor receptors such as epidermal growth factor receptor (EGFR) and c-KIT

(Nalwoga *et al.*, 2008; Nielsen *et al.*, 2004; Sorlie *et al.*, 2003; Sotiriou *et al.*, 2003). In contrast, the luminal types of tumors (luminal A and luminal B) generally express both the ER and PR as well as cytokeratins associated with mammary epithelial cells of the luminal lineage (Perou *et al.*, 2000; Rakha *et al.*, 2007b). Patients with luminal A tumors have a better prognosis than patients with luminal B tumors, and these tumors can generally be discriminated from one another on the basis of tumor grade, where luminal B tumors are of high grade, similar to basal-like and Her2+ tumors (Loi *et al.*, 2007; Parker *et al.*, 2009; Sorlie *et al.*, 2001). The HER2 overexpressing subgroup is characterized by the overexpression of HER2 and other genes on the 17q amplicon, such as *GRB7,* however HER2 over-expression has also been observed in some luminal B tumors (Sorlie *et al.*, 2003). As gene expression studies continued to evolve, new molecular subtypes of breast cancer continue to be discovered; for example a claudin-low subtype has recently been identified (Herschkowitz *et al.*, 2007; Prat *et al.*, 2010).

These aforementioned molecular subtypes also display substantial genetic differences. For example, basal-like breast tumors possess a high level of gene copy number variation relative to the other subtypes, suggesting that basal-like breast tumors have increased genomic instability (Chin *et al.*, 2006; Chin *et al.*, 2007; Holstege *et al.*; Kwei *et al.*). Basal-like tumors are typically enriched in low-level copy-number gains at multiple loci, whereas high-level amplification of a particular locus is rare. Conversely, in luminal and Her2+ tumors, a smaller number of high level amplifications are much more frequent (Holland *et al.*, 2011; Sircoulomb *et al.*, 2011; Staaf *et al.*, 2010).

Although substantial research has focused on studying the molecular subtypes of breast cancer, targeted therapies have yet to emerge as a result of the classification scheme. From a clinical perspective, the contribution of the molecular subtypes has largely been the observation that patients with luminal A tumors have an excellent prognosis and derive little, if any, benefit from adjuvant chemotherapy, whereas patients with tumors of the other subtypes have a poorer prognosis and also more likely to derive benefit from intervention with adjuvant chemotherapy (Hatzis *et al.*, 2011; Rouzier *et al.*, 2005a). A classification scheme for the molecular subtypes of breast cancer has been developed clinically, although it has not been tested for utility in a prospective setting (Parker *et al.*, 2009).

Gene expression signatures and patient prognosis

In concert with the identification of molecular subtypes of breast cancer, many studies focused on the development of multigene predictors that could be used to guide breast cancer patient prognosis (Paik *et al.*, 2004; van 't Veer *et al.*, 2002; van de Vijver *et al.*, 2002). Whereas over the last decade many prognostic signatures have been described and published, such as a 21-gene (Paik *et al.*, 2004), 70-gene(van 't Veer *et al.*, 2002), 76-gene (Wang *et al.*, 2005), 77-gene genomic grade profile (Sotiriou *et al.*, 2006), wound response signature (Chang *et al.*, 2005) and others (Hallett *et al.*, 2010; Hallett and Hassell, 2011; Rody *et al.*, 2009; Yau *et al.*, 2010), relatively few have been implemented in clinical practice (Colombo *et al.*, 2011; Sotiriou and Pusztai, 2009; Weigelt *et al.*, 2010). Moreover, the prognostic power of these signatures appears to be

derived from their capacity to measure proliferation and ER signaling (Desmedt *et al.*, 2008; Wirapati *et al.*, 2008). As a result, these signatures typically classify ER- and high grade ER+ tumors as high risk, and therefore have negligible prognostic value for ER- patients. Some recent work has focused on identifying multi-gene predictors of outcome in triple negative and ER-/PR- receptor negative breast cancers (Desmedt *et al.*, 2008; Hallett *et al.*, 2012; Kreike *et al.*, 2007; Rody *et al.*, 2009; Sabatier *et al.*; Teschendorff *et al.*, 2007; Yau *et al.*, 2010), however these findings still require validation in additional datasets. Two prognostic gene signatures, Mammaprint® (Agendia) and Oncotype DX® (Genomic Health) are currently being tested in prospective phase III clinical trials to evaluate their use in clinical practice.

Mammaprint® is a 70-gene assay developed from a retrospective series of 78 node-negative breast cancer patients who did not receive adjuvant chemotherapy (van 't Veer *et al.*, 2002). The expression of the 70 genes is used to assign patients into a good or poor risk group, and has been approved by the FDA to guide the prognosis of breast cancer patients (Sotiriou and Pusztai, 2009). Importantly, this assay has been validated retrospectively in a 307 patient cohort who did not receive adjuvant chemotherapy (Buyse *et al.*, 2006). Mammaprint® risk assignment was also compared with risk assignment based on conventional criteria, such as tumor size, grade, and nodal status, using the Adjuvant! Online program (http://www.adjuvantonline.com/index.jsp). In approximately 30% of cases, Mammaprint® and Adjuvant! Online produced discordant results and in these cases Mammaprint® risk assignment appeared to be more accurate. Clinically high-risk patients that were low-risk by the Mammaprint assay experienced an

excellent 10-year survival rate of ~90%, whereas survival in patients with clinically low-risk tumors but high risk by the Mammaprint® assay was a much poorer ~70%. In a multicenter prospective study, the use of Mammaprint® led to recommendations for altered treatment in ~25% of patients (Bueno-de-Mesquita *et al.*, 2007). MINDACT, a large phase III randomized clinical, is currently ongoing to evaluate widespread use of Mammaprint® to guide the use of therapy in breast cancer patients (Rutgers *et al.*, 2011).

Oncotype DX® (Genomic Health) combines the expression of 21 genes into a quantitative recurrence score (RS) that can be used to estimate risk recurrence after 10-years (Paik *et al.*, 2004). The genes were selected using a candidate gene approach from an initial list of 250 genes tested for association with survival in a relatively large training cohort. The RS is a continuous variable from 0-100, which is used to define 3 risk categories, low (RS<18), intermediate (18<RS<31), and high (RS=>31). Oncotype DX® has been validated in a relatively large cohort of ER+, tamoxifen-treated, node-negative breast cancer patients (n=668) enrolled in the NSABP B-14 trial, which showed that the rate of distant recurrence was ~7%, ~14% and ~31% for low, intermediate, and high risk patients, respectively (Paik *et al.*, 2004). Similar results were also reported in a similar community based cohort (Habel *et al.*, 2006). In both cases, the association between RS and survival was independent of standard clinical pathological variables, confirming that the RS can add important information when determining patient prognosis. Importantly, Oncotype DX® can also identify tumors, which are more likely respond to chemotherapy. In a cohort of 651 patients enrolled in the NSABP B-20 clinical trial, higher RS were associated with benefit from adjuvant chemotherapy, whereas lower RS

scores were associated with negligible benefit (Paik *et al.*, 2006). Oncotype DX® has been endorsed by the American Society of Clinical Oncologists (ASCO) as useful to inform prognosis in patients with ER+ and node negative tumors (Harris *et al.*, 2007).

Gene expression signatures and patient response to chemotherapy

Whereas the implementation of gene signatures that guide patient prognosis into the clinic has been relatively successful, similar signatures that seek to predict response to different chemotherapy regimens have yet to yield clinically useful tests. The goal of such predictive signatures is to select the most appropriate regimen for a given patient, to achieve the greatest pathological response. Perhaps the best example of therapy based patient selection is illustrated by the example of trastuzumab, where in the absence of patient selection, trastuzumab is clinically beneficial in slightly less than 10% of breast cancer patients. However, when patients are selected for trastuzumab therapy based on *HER*2 amplification, the response rate rises dramatically to 35-50% (Vogel *et al.*, 2002).

Measuring the efficacy of chemotherapy in a tumor can be accomplished by assessment of response following neoadjuvant treatment with a given treatment regimen (Symmans *et al.*, 2007). In this fashion, patients with little to no invasive or metastatic disease detected following treatment are classified as having minimal residual cancer burden (RCB0/RCBI), whereas patient's whose tumors show less response and are classified as having greater residual cancer burden (RCBII/RCBIII). Importantly, neoadjuvant chemotherapy has been found to be as efficacious as chemotherapy given in

the adjuvant setting, and patients who achieve complete pathological response after neoadjuvant intervention generally have an excellent probability of experiencing long-term survival (Bear *et al.*, 2006; Liedtke *et al.*, 2008; Wolmark *et al.*, 2001). Taken together, these data suggest that response to neoadjuvant chemotherapy (RCB0/I/II/III) is a relevant clinical model to develop and validate gene signature based predictors of breast tumor response to chemotherapy. Archival samples, which could be used empirically to train and validate gene signatures that predict regimen specific responses are rare, requiring that such specimens be collected in a prospective manner. However, in some cases researchers have taken such an empirical approach with relatively few samples for gene signature discovery and validation (Ayers *et al.*, 2004; Hess *et al.*, 2006), although it is unclear whether these signatures are independently predictive of response to chemotherapy when combined with standard clinical variables (Lee *et al.*; Rouzier *et al.*, 2005b). Alternative methods have focused on using the responses of experimental cell line models to therapies as a means to identify genes whose expression is predictive of agent-specific responses. Although the initial results produced using this methodology seemed quite promising (Potti *et al.*, 2006; Potti and Nevins, 2008), many of these seminal studies could not be reproduced and have since been retracted (Borst and Wessels, 2010; Liedtke *et al.*, 2010; Potti *et al.*, 2006; Potti and Nevins, 2008). Differences in response rates to chemotherapy have also been observed among the different molecular subtypes of breast cancer, where early studies suggest they are approximately ~7% for luminal A, 17% for luminal B, 36% for HER2+, and 43% for

basal-like tumors (Parker *et al.*, 2009). Hence, the classification of tumors into these subtypes may also be useful to predict response to chemotherapy.

Empirical approaches, such as the ones described above, attempt to identify predictive gene signatures by comparing the gene expression profiles of responsive tumors to non-responsive tumors. This methodology assumes that mechanisms of resistance or sensitivity are consistent between different primary tumors, and that this common difference will be detectable using gene expression profiling. There is increasing experimental evidence that drug resistance can be achieved through multiple distinct mechanisms. For example, multiple studies suggest that resistance to trastuzumab can be achieved through downstream mutation of the tumor-suppressor gene PTEN(Nagata *et al.*, 2004), expression of a truncated HER2 protein (Scaltriti *et al.*, 2007), or overexpression of either IGFR (Lu *et al.*, 2004; Lu *et al.*, 2001), HER3(Lee-Hoeflich *et al.*, 2008), or MUC4(Carraway *et al.*, 2001). Indeed, these observations suggest that the identification and development of predictive gene signatures to guide selection of chemotherapy regimens for breast cancer patients will be very difficult.

Novel therapies: Targeting cancer stem cells

Many of the studies discussed above aim to optimize the use of approved therapies in breast cancer patients. However, even with excellent patient management, currently approved therapies may fail to provide durable cure and these patients relapse and succumb to their disease. Many studies have identified subpopulations of cells within tumors, termed cancer stem cells or tumor-initiating cells (TIC), which are the

suspected culprits responsible for driving tumor relapse (Al-Hajj *et al.*, 2003; Clevers; Collins *et al.*, 2005; Fang *et al.*, 2005; Lapidot *et al.*, 1994; Lobo *et al.*, 2007; O'Brien *et al.*, 2007; Prince *et al.*, 2007; Reya *et al.*, 2001; Singh *et al.*, 2003). Accordingly, many studies are currently focused on identifying additional biomarkers of breast TICs as well as novel therapeutics that target these cells.

The cancer stem cell hypothesis posits that tumors comprise a cellular hierarchy, which is a caricature of their normal tissue counterpart, and that tumor growth is driven by rare TICs that sit at the apex of the tumor cell hierarchy (Lobo *et al.*, 2007; Spillane and Henderson, 2007). Several important observations led to the development of this hypothesis. One of these is that although most tumors arise clonally from a single cell, tumors consist of a heterogeneous cell population. Cancer cells within tumors comprise a diversity of morphologies, immunophenotypes, and proliferative potentials (Heppner, 1984). Another observation was that a large number of tumor cells were required for transplant to seed the growth of secondary tumors (Bruce and Van Der Gaag, 1963). The explanation for these observations, provided by the cancer stem cell hypothesis is that only the rare TICs can initiate tumor growth. Due to the scarcity of these cells, only transplants of a sufficient number of tumor cells would ensure that a requisite TIC, necessary to drive tumor formation and growth, was included in the cell transplant population. Because TICs also differentiate, albeit aberrantly, this causes them to give rise to multiple cell types, which are reminiscent of their normal tissue counterparts, accounting for the diversity of cell types observed within tumors (Lobo *et al.*, 2007). The existence of TICs was proven by fractionating tumor cells into different groups based on

their expression of cell surface markers, and transplanting these various fractions into mice. Only some of the tumor cell fractions were capable of engrafting and eliciting tumor growth in mice, whereas others could not, even when large numbers of cells were transplanted. These experiments provided the foundation on which the theory of cancer stem cells, or TICs, was founded.

Tumor-initiating cells

Although the notion of TICs has existed for almost half a century (Bruce and Van Der Gaag, 1963; Hamburger and Salmon, 1977; Pierce and Johnson, 1971) it wasn't until 1994 that John Dick's group provided strong evidence for the existence of TICs (Lapidot *et al.*, 1994). Their experiments relied on the use of the severe combined immunodeficiency disease (SCID) immunodeficient mouse model as an experimental system in which they could measure the growth of transplanted human acute myeologeneous leukemic (AML) cells. By using fluorescent activated cell sorting (FACS) and fluorescently tagged antibodies, which recognized, cell surface proteins, they were able to separate the AML cells into tumorigenic and non-tumorigenic cell populations. Specifically, only the CD34+/CD38- population, indicating that the surface of these cells expressed CD34 but not CD38, were able to give rise to leukemia in the SCID mice, demonstrating that the CD34+/CD38- cell fraction contained the TICs (Lapidot *et al.*, 1994). This strategy is now commonly used to isolate and identify functional TIC, and has been used to find TICs in other human tumor types. To date, TICs have also been identified in tumors of the breast (Al-Hajj *et al.*, 2003), brain (Singh

et al., 2003), skin (Fang *et al.*, 2005) , gastro-intestinal tract (Haraguchi *et al.*, 2006; O'Brien *et al.*, 2007), prostate (Collins *et al.*, 2005), and head and neck (Prince *et al.*, 2007). Importantly, these discoveries suggest that TICs underlie a large proportion, if not all malignancies, and are not an idiosyncracy of leukemia.

Clinical Implications of Cancer Stem Cells

The discovery of TICs has important implications for cancer therapy. Namely, that cancer treatments need to eliminate the TICs from tumors in order to be both effective and durable. Most current cancer therapeutics were discovered based on their ability to kill human tumor cell lines and to shrink xenografts in mice that were seeded by these same lines (Shoemaker, 2006). These cell lines and their xenograft counterparts comprise rapidly proliferating tumor cells and as a result most chemotherapies are cytotoxic anti-mitotic therapies. TICs, like normal adult stem cells, are generally though to divide relatively infrequently, and can therefore avoid the killing effect of these anti-mitotic cancer drugs. Furthermore, TICs have also been found to have both highly efficient drug export mechanisms (Eramo *et al.*, 2006; Ghods *et al.*, 2007; Kang and Kang, 2007; Ma *et al.*, 2008; Wulf *et al.*, 2001), and heightened resistance to radiation therapy (Bao *et al.*, 2006; Baumann *et al.*, 2008), properties which allow them to survive the killing effects of these anti-cancer therapies; hence TICs are likely not killed by currently used cancer therapeutics, and account for cancer reccurence in cancer patients. The key to providing long term cancer cure is to find a means to destroy or abrogate the tumorigenicity of TICs, thus eliminating the possibility of tumor reccurrence.

The expression profiles of TICs have also been used to develop prognostic gene expression signatures to predict cancer survivorship and metastasis. The greater the similarity between a patient's tumor's gene expression profile and the TIC gene signature, the higher the risk of relapse for that patient (Eppert *et al.*, 2011; Liu *et al.*, 2007b). Importantly, one such signature derived using breast TICs, but was also prognostic in other types of cancer (Liu *et al.*, 2007b). Not only do these results suggest that the frequency of TICs in a tumor may be related to the outcome of the cancer patient, but that TICs of different tumor types may also share common characteristics. These experiments were important because they were the first to suggest that TIC biology is linked to outcome in human patients, further corroborating that TICs represent the most important cellular targets for future anti-cancer therapies.

Breast tumor-initiating cells

In 2003 Michael Clarke's group discovered TICs in human breast tumors (Al-Hajj *et al.*, 2003). By building on John Dick's work in AML, they were able to fractionate human breast tumor cells into tumorigenic and non-tumorigenic cell populations based on their expression of cell surface markers. They found that as few as 100 of the tumorigenic cells phenotypically defined as $CD44^{+}CD24^{-/low}Lin^{-}$ form tumors after orthotopic transplant into mice, whereas several tens of thousands of cells of other phenotypes, such as $CD44^{+}CD24^{high}$, could not. Furthermore, serial transplantation of the secondary tumors confirmed that the TICs were able to self-renew as well as give rise to the phenotypically diverse non-tumorigenic cells that recapitulated the cellular

composition of the original tumor (Al-Hajj *et al.*, 2003). More recent research has discovered other markers of breast TIC. Human breast tumor cells can also be sorted into TIC and non-TIC fractions using the markers ALDH1 or GD2 (Battula *et al.*, 2012; Ginestier *et al.*, 2007). These markers can also be used to further fractionate the $CD44^{+}CD24^{-/low}Lin^{-}$ CSC enriched population, and in the case of ALDH1 tumors can be generated from injecting as few as 20 $ALDH1^{+}/CD44^{+}CD24^{-/low}Lin^{-}$ cells. The ALDH1 cell phenotype in human breast tumors was also found to correlate with clinical outcome and histoclinical grade (Ginestier *et al.*, 2007).

Some studies suggest that induction of an epithelial-to-mesenchymal (EMT) transition in either normal or transformed epithelial cells results in their acquisition of stem cell properties (Mani *et al.*, 2008). Moreover, induction of EMT by forced expression of either Snail or Twist, or exposure to TGFβ in immortalized human mammary epithelial cells increased both the fraction of cells that were CD44+/CD24-, as well as their capacity to form mammospheres and initiate tumor growth (Gupta *et al.*, 2009). Importantly, these data suggest that breast TICs may exhibit some level of plasticity and may exist in a phenotypic equilibrium with non-TICs. This notion is supported by recent observations that stochastic state transitions occur in cancer cell populations and produce an equilibrium of TIC and non TIC cells (Gupta *et al.*, 2011).

These various experiments described above provide evidence that breast cancer follows a cancer stem cell or TIC model, where tumor growth is driven by rare TICs and that the aberrant differentiation of their progeny is the cause of tumor cell heterogeneity.

However, these various observations also suggest that the cancer stem cell/TIC model may be less straightforward than previously envisioned, and that stochastic transition from a non-TIC to a TIC state may occur at low frequencies. From a clinical perspective, it is obvious that targeting TICs will be necessary to develop more effective and durable breast cancer therapies. Accordingly, studies have reported identification of agents that target TICs, but few, if any, have begun testing in early stage clinical trials (Gupta *et al.*, 2009; Sachlos *et al.*, 2012). Detailed analyses of these cells have been hampered largely as a result of their scarcity. Even when using the most advanced cell sorting techniques, it is only possible to fractionate breast tumor cells to 5% TIC purity (Al-Hajj *et al.*, 2003; Ginestier *et al.*, 2007). As an alternative, researchers have turned to other models of cancer stem cells in order to expedite their studies.

Mouse Models of Breast Cancer Stem Cells

To overcome the limitations of studying TICs in human breast tumors, many researchers tried to identify TICs in mouse models of breast cancer. In 2007, Eldad Zacksenhaus group in Toronto were able to fractionate mammary tumor cells from the MMTV-*Neu* model based on the expression of CD24(Liu *et al.*, 2007a). Transplant of $CD24^{high}$ but not $CD24^{low}$ cells was able to engraft and form tumors in mice. Furthermore, the tumors that arose from transplanted $CD24^{high}$ cells phenocopied the original primary tumor, indicating that these cells were not only able to self-renew, but also to differentiate into non-tumorigenic $CD24^{low}$ cells. Similar studies by Michael Clarke's and Jeffrey Rosen's groups, with the MMTV-*Wnt1* and p53-null mouse

mammary tumor models, have also uncovered TICs (Cho *et al.*, 2008; Zhang *et al.*, 2008). As is the case with human breast tumors, the frequency of TICs in these mouse mammary tumors is still exceedingly low (Cho *et al.*, 2008; Liu *et al.*, 2007a; Zhang *et al.*, 2008). Even after enriching for the TICs by fractionating the tumor cells based on their expression of cell surface marker proteins, TICs generally only comprise 1-2% of the partially purified cell population. In contrast to the work of other labs, we have found that tumors from the MMTV-*Neu* [N202], MMTV-Δ*Nβ-catenin, and* MMTV-*PyMT* mouse models of breast cancer can comprise a major TIC cell fraction, typically 30%-90% of the total cell population. The rich source of TICs afforded by these mouse models have allowed us to begin comprehensive analyses of these cells.

Rationale and Objectives

Despite significant research, breast cancer remains a leading cause of cancer mortality among women worldwide (Jemal *et al.*, 2011). Advances in technology that enable high-throughput gene expression profiling, such as microarrays and quantitative RT-PCR platforms, have led to the use of molecular signatures to discern important biological information about breast cancer. The main objective of my research is to use gene expression profiling technologies to probe multiple aspects of breast cancer biology, which will be described in detail over the following chapters. Chapters 2-4 comprise complete manuscripts that describe gene signatures that are designed to serve as biomarkers to guide the use of approved and experimental therapies in breast cancer patients. Chapters 5 and 6 describe gene expression signatures that represent breast TIC

biology and were utilized to identify agents that target these cells. Ultimately, we hope to use the knowledge garnered throughout this research to more effectively deliver approved therapies to breast cancer patients, as well as identify novel therapies that benefit those patients for whom available treatments fail to provide durable cancer cure.

CHAPTER 2

Introduction

Traditionally a number of tumor characteristics have been used to determine the prognosis of breast cancer patients. Such factors include tumor size, grade, hormone receptor status, HER2 status, lympho-vascular space invasion and lymph node involvement (1998; Hayes *et al.*, 1998). More recently whole genome analysis technology (gene expression profiling) has been added to the armamentarium of experimental techniques, thus providing a new molecular classification for breast cancer and contributing to the development of a number of prognostic multi-gene assays as well as the identification of the molecular subtypes of breast cancer. Although initial experiments suggested that basal-like breast cancers (BLBC) were associated with a particularly poor prognosis (Sorlie *et al.*, 2001; Sorlie *et al.*, 2003), however on closer examination these studies demonstrated that the prognosis of patients with BLBCs is highly time dependent. Some patients with BLBCs experience particularly poor survival in the first 3-5 years following diagnosis, but for others their mortality wanes such that at 10 years post diagnosis these patients have a better survival than those with luminal-type (ER+) tumors (Blows *et al.*; Cheang *et al.*, 2008; Dent *et al.*, 2007; Mulligan *et al.*, 2008). This suggests that patients with BLBCs can be separated into two clinically distinct groups: those likely to experience a recurrence and to succumb to their disease in the first 3-5 years after diagnosis, and those expected to show excellent long term survival.

Whereas several multi-gene signatures exist to predict breast cancer patient prognosis, their prognostic values appears to be mostly derived from their capacity to

measure expression of genes associated with proliferation (Desmedt *et al.*, 2008; Wirapati *et al.*, 2008). Because BLBCs are generally highly proliferative, the existing prognostic signatures fail to identify a subset of BLBC with good prognosis (Yau *et al.*). Some recent work has focused on identifying multi-gene predictors of outcome in triple negative (ER-, PR-, HER2-) and hormone-receptor negative breast cancer (Desmedt *et al.*, 2008; Karn *et al.*; Kreike *et al.*, 2007; Rody *et al.*, 2009; Rody *et al.*; Sabatier *et al.*; Teschendorff *et al.*, 2007; Yau *et al.*). However, a robust method of distinguishing between BLBCs with good and poor outcome has yet to be developed.

In this manuscript we report the identification and validation of a 14-probe set signature that can stratify BLBCs into high and low risk groups. Whereas the data described here comprises experimental data derived using microarrays, we have extended these findings in ~90 clinical samples obtained from Hamilton Health Sciences using Nanostring's nCounter Analysis System™, and are currently trying to 'accrue' more samples for additional validation. Ultimately, we'd like to complete a prospective clinical trial to justify the use of our signature to guide the use of chemotherapy in BLBC patients.

Manuscript #1

Title: A Gene Signature for Predicting Outcome in Patients with Basal-like Breast Cancer

Robin M. Hallett[1], Anna Dvorkin-Gheva[1], Anita Bane[2] and John A. Hassell[1,*]

1. Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5
2. Department of Oncology, McMaster University, Juravinski Cancer Centre, 699 Concession St East, Hamilton, Ontario, Canada, L8V 5C2

*Correspondence and requests for materials should be addressed to J.A.H. (hassell@mcmaster.ca)

Acknowledgements

This work was generously supported by grants from the Canadian Stem Cell Network, the Ontario Institute for Cancer Research and the Canadian Breast Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Robin Hallett – Conceived, planned, analyzed and performed most of the experiments in the paper, wrote the manuscript

Anna Dvorkin-Gheva – Generated data for figure 1a, wrote section of methods

Anita Bane – Involved in conception of project, helped write manuscript, provided critical feedback for project

John Hassell – Helped write manuscript, provided critical feedback for project

Abstract:

Basal-like breast cancer is a molecular subtype of breast cancer with a poor prognosis. Follow-up studies of long-term outcome in these patients, demonstrates they can be separated into two clinical groups: those who succumb to their disease within the first 5 years and those expected to show excellent long term survival. Currently available clinical/histopathological variables as well as molecular signatures show little capacity to identify basal breast cancer patients with either a high or low risk of disease relapse. Using data derived from 85 basal-like breast cancer patients, we identified a 14-gene signature, which we subsequently validated on an additional 49 basal breast cancer patient set. The ability to distinguish between these two sub-groups of basal breast cancer patients at the time of initial diagnosis would permit tailoring aggressive therapeutic regimens to those patients with a poor prognosis and conversely avoid such therapy in low risk patients.

Results

Compiling multiple gene expression profiles of basal breast tumors

To identify genes whose expression might be associated with the clinical outcome of BLBC patients, we compiled a large collection of human breast tumor gene expression data for which clinical data was also available (n=995). Hierarchical clustering using the 'intrinsic' gene set revealed that many of these tumors (n=547) clustered into the previously described molecular subtypes (Perou *et al.*, 2000; Sorlie *et al.*, 2001; Sorlie *et al.*, 2003) (Fig. 1a). Importantly, survival analysis using Kaplan-Meier survival curves revealed distinct differences in clinical outcome among the patients with tumors of different molecular subtypes. As observed previously patients with tumors of the basal-like, ERBB2, claudin-low and luminal B subtypes experienced the poorest 10-year survival, whereas patients with luminal A or normal-like tumors experienced the best 10-year survival (Sorlie *et al.*, 2001) (Fig. 1b). Interestingly, the 10-year survival rate of patients with basal-like tumors was approximately 60% and very few BLBC patient mortalities occurred after this time (Fig. 1b). The latter findings are consistent with previous observations that the prognosis of BLBC patients is time dependent, where these patients are at highest risk for relapse during the first 5 years post diagnosis and experience a very low risk for relapse 10-years post diagnosis (Cheang *et al.*, 2008; Dent *et al.*, 2007; Mulligan *et al.*, 2008).

Importantly, the BLBC tumor cohort comprised 134 patients with clinical follow-up data, thus providing a fairly large number of basal tumors to identify a genomic predictor that could be used to guide prognosis for patients with basal-like breast tumors.

Training signature

To develop a genomic predictor that could be used to identify BLBC patients who were likely to have either good or poor survival outcomes, we first divided the 134 patient BLBC cohort into a 85 patient training set and a 49 patient validation set. We used binary regression probabilistic models for feature selection to identify genes that had the best prognostic performance among the gene expression profiles derived from the 85 BLBCs of the training set (West *et al.*, 2001). For these analysis, <5 year DFS was taken to indicate poor outcome, whereas >5 year DFS was taken to indicate good outcome. Previous studies have shown that the vast majority of disease recurrence among BLBC patients occurs within the first 5 years (Cheang *et al.*, 2008; Dent *et al.*, 2007; Mulligan *et al.*, 2008). Starting with a single probe set signature, we iteratively generated signatures by gradually adding probe sets and tested the resulting signature using leave-one-out cross-validation. In this fashion we generated multiple signatures comprising n probe sets, where n = 1,2,3…,50 (Fig2. A). For each discrete value of n, this technique assigned a probability to every patient within the training set that indicated the likelihood of a patient experiencing disease relapse. To establish a probability cut-point, where patients with higher probability are assigned into the poor prognosis category and patients with lower probability are assigned into the good prognosis category, we used a previously

described tertile method (Haibe-Kains *et al.*, 2008). In this fashion, good prognosis was assigned to patients whose probability score fell in the lowest 1/3 of all probability scores, whereas poor prognosis was assigned to patients whose score fell into the higher 2/3 of probability scores. Indeed, these approximate proportions have been observed in several gene expression based breast cancer prognostication studies(Buyse *et al.*, 2006; Desmedt *et al.*, 2007; van 't Veer *et al.*, 2002; Wang *et al.*, 2005). We therefore took this approach as a relatively non-biased and simple means to divide patients into predicted good and poor outcome groups. To determine which n-element signature had optimal performance we compared the relative risk of relapse for each signature (Fig. 2B, red line: relative risk, black line: LOWESS (LOcally WEighted Scatterplot Smoothing) curve fitted to relative risk data, n=14 identifies optimal signature length). In this fashion we identified a 14-probe-set (each gene represented by 1 probe set, henceforth called Basal 14 signature) signature, which optimally separated patients into good and poor outcome groups (Table 1).

Assessment of Signature Performance

Validation of a gene signature using an independent data set is a more accurate measurement of its prognostic value than using cross-validation on a training data set. Therefore, we tested our Basal-14 signature on an independent cohort of patients with BLBC (n=49). To learn whether the probability of disease relapse predicted by the Basal-14 signature could be used as a continuous predictor of disease relapse, we calculated the proportion of patients who had experienced disease relapse while increasing the cut-off

(decreasing stringency) for assigning a patient into the good outcome group. Indeed, the proportion of patients experiencing disease relapse increased in an approximate linear fashion as the probability assigned for disease relapse by the Basal-14 signature increased (Fig. 3A). To assess the predictive accuracy of the Basal-14 signature, we completed receiver-operator characteristic (ROC) curve analysis. In this fashion, an AUC (Area Under Curve) value of 0.5 indicates predictive performance which is no better than chance, whereas values greater than 0.5 indicate true predictive capacity. The Basal-14 signature produced an AUC that was statistically significantly higher than 0.5 (AUC: 0.76, p=0.003, Fig. 3B). Taken together, these data demonstrate the capacity for the Basal-14 signature to identify BLBC patients at high risk for disease relapse. To visualize survival differences between groups of patients that were predicted to have either high or low risk for disease relapse, we stratified patients from the validation cohort into good and poor outcome groups using tertiles, and completed Kaplan-Meier survival analysis. Patients whose predicted probability for disease relapse fell within the lowest tertile of predicted probabilities were stratified into the good outcome group, whereas those whose predicted probabilities fell within the upper two tertiles were stratified into the poor outcome group. The Kaplan-Meier estimate for the proportion of patients in the low-risk group who did not experience a disease relapse at 5 years (94%) was significantly greater than the proportion in the poor outcome category (48%) (Table 2, Fig. 3C, HR: 4.7 [CI95: 1.8-12.3], p = 0.0017). Because our overarching objective was to identify patients who could be spared aggressive chemotherapy, we also tested the capacity of our signature to predict the outcome of patients who had not received adjuvant chemotherapy. In this

fashion, we were able to test the relationship between the Basal-14 signature and the natural progression of BLBCs without having adjuvant chemotherapy as a potentially confounding variable. 26 patients within the 49 patient validation cohort met this criterion (patients from GSE7390 & GSE2034). We re-tested the predictive capacity of the Basal-14 signature on these 26 chemotherapy naïve patients and observed a statistically significant difference in the survival of patients who were predicted to have either good or poor outcome (Fig. 3D, HR: 4.4 [CI95: 1.1-16.7], p = 0.03, Table 3). The proportion of patients in the chemotherapy naïve validation cohort who were predicted to have good survival and were free of disease at 5 years was 100%, whereas among those patients who were predicted to have poor survival, only 50% were disease free after 5 years. Taken together, these findings demonstrate the capacity of our gene signature to identify patients who have excellent long-term survival even when patients did not receive aggressive adjuvant chemotherapy.

Comparison of the Basal-14 signature with other multigene predictors

Previous studies have reported that many published multigene predictors fail to accurately identify high and low risk patients among patients with ER-negative breast cancer (Teschendorff *et al.*, 2007; Yau *et al.*). As the majority of BLBCs are ER-negative, we sought to test whether multiple previously described multigene predictors were prognostic in the context of BLBC. To this end, we measured the association of the Genomic Grade Index (Sotiriou *et al.*, 2006), NKI-70 signature (van de Vijver *et al.*, 2002), Recurrence score (Paik *et al.*, 2004), CSR/Wound response signature (Chang *et*

al., 2005), Triple-negative signature (Yau *et al.*), MS-14 signature (Tutt *et al.*, 2008), as well as the Basal-14 signature in the 49 patient validation cohort by calculating a signature index and completing either Kaplan-Meier survival analysis using tertiles to dichotomize the validation cohort into good and poor outcome groups, or generating ROC curves. Interestingly, other than the Basal-14 signature (Fig. 4A, HR: 4.3 [CI95: 1.6-11.4], p = 0.0032) none of the other signatures identified patient groups with statistically significant differences in survival (Kaplan-Meier: Fig. 4A-F. ROC: supplementary figure 1A-F). These data suggest that the prognostic capacity of previously reported multigene outcome predictors may be diminished in patients with BLBC. However, it should be noted that the tertile method used to separate patients into good and poor outcome groups may be non-optimal for these signatures. Interestingly, the triple negative signature trended towards significance in the Kaplan-Meier analysis (Fig. 4F, HR: 2.0 [CI95: 0.8-5.4], p = 0.15) and was statistically significant in the ROC curve analysis (Supplementary fig. 1G, AUC: 0.7, p = 0.02). This is likely because the triple negative signature was developed with triple negative breast tumors, which comprises a sub-group that overlaps with the basal-like molecular subtype. Together, these findings underscore the need for prognostic multigene signatures, such as the Basal 14 signature, for guiding therapy choice for breast cancer patients.

Performance of Basal-14 signature in other molecular subtypes of breast cancer

Previous studies have demonstrated that biological processes that can be linked to breast cancer patient outcome vary among the different molecular subtypes of breast

cancer (Desmedt *et al.*, 2008). In this regard, we sought to test whether the Basal-14 signature could be used to identify high and low risk patients among the other molecular subtypes of breast cancer, or whether its capacity to stratify patients into high and low risks groups was limited to patients with BLBCs. The Basal-14 signature showed no capacity to identify patients at high and low risk for disease relapse among the luminal A (HR: 1.3, p=n.s.), luminal B (HR: 1.2, p=n.s.), claudin low (HR: 1.0, p=n.s.) and normal (HR: 0.4, p=n.s.) molecular subtypes of breast cancer (Fig. 5A-D). Unexpectedly, the Basal-14 signature was also prognostic in the ERBB2 molecular subtype (HR: 2.8 [CI95: 1.3-6.5], p = 0.01). Interestingly, a previously reported prognostic gene signature developed using Her2-positive tumors was also found to be prognostic in BLBCs (Staaf *et al.*). These data suggest that similar biological processes may govern patient outcome in both the basal-like and ERBB2 molecular subtypes of breast cancer. Taken with our previous findings, it appears that transcripts whose expression may be informative for patient prognosis vary between the different molecular subtypes of breast cancer. For example, it appears that signatures that are prognostic in ER-positive breast tumors, such as the Reccurrence score (OncotypeDX®) and the Genomic Grade Index, fail to stratify BLBCs into good and poor outcome groups, whereas the Basal-14 signature is prognostic in basal-like and ERBB2-overexpressing breast cancer, but fails to identify patients in the ER-positive luminal subtypes of breast cancer.

Discussion

Few, if any, clinical variables show prognostic capacity in the context of BLBC. Therefore, we sought to identify a genomic predictor of patient outcome for patients with BLBC. In the present study, we identified a 14 probe set signature, which we named the Basal 14 signature. We tested the Basal 14 signature on an independent validation cohort of BLBC patients and were able to accurately stratify patients into good and poor outcome groups. Importantly, the difference in risk for disease relapse for patients who were predicted to have either good or poor outcome was both relatively large and statistically significant. Because it was unclear whether the Basal 14 signature was related to the natural progression of BLBCs, tumor response to therapy, or both, we also tested the Basal 14 signature on a smaller group of patients who did not received treatment with adjuvant chemotherapy. In this fashion, we were able to confirm a relationship, albeit in a small number of patients, between the Basal 14 signature and patient survival in chemotherapy naïve patients. Notably, previous reports suggest that immune-based signatures predict response to chemotherapy in triple negative breast cancer patients, suggesting that the Basal 14 signature might also measure treatment response (Desmedt *et al.*, 2008; Hess *et al.*, 2006). The relationship between the Basal 14 signature and response to chemotherapy was not examined in this study. Another possibility is that the Basal 14 signature is associated with histological subtypes of BLBC with known good prognosis, such as the medullary subtype(Ridolfi *et al.*, 1977; Vincent-Salomon *et al.*, 2007). However, the frequency of medullary breast tumors is exceptionally low (2%), suggesting that the Basal 14 signature would also need to identify good prognosis non-

medullary BLBCs to achieve the level of accuracy described here. In total, the capacity of the Basal 14 signature to identify BLBC patients with good prognosis is likely multi-factorial, and many additional possibilities remain unexplored.

Interestingly, the Basal 14 signature comprised multiple genes with known roles in cancer. For example, destrin (DSTN) is one of three mammalian actin depolymerisation factors (ADFs). These proteins are fundamental for multiple cellular processes such as cell survival, cytokinesis, as well as cell migration and chemotaxis (Van Troys *et al.*, 2008), and have been linked as a major determinant of metastasis in cancer patients (Wang *et al.*, 2007; Wang *et al.*, 2004). Tudor domain containing protein 3 (TDRD3) has previously been linked to outcome in patients with ER-negative breast tumors (Nagahata *et al.*, 2004), and while being relatively poorly characterized, is thought to play a role in the regulation of cytoplasmic stress granules (Goulet *et al.*, 2008). Regulator of G-protein signaling (RGS4) has also been linked to patient outcome in patients with triple negative tumors (Yau *et al.*). Notably, RGS4 appears to be a key negative regulator of breast cancer cell migration and invasion (Xie *et al.*, 2009). It is therefore somewhat surprising that high levels of RGS4 transcripts are associated with poor outcome. However, it appears that RGS4 function is heavily regulated post-translationally by proteosomal degradtaion, suggesting that a negative feedback loop occurs where high levels of RGS4 transcripts indicate low levels of RGS4 protein (Xie *et al.*, 2009). Interestingly, proteasome inhibitors are being explored as possible means for cancer therapy (Orlowski and Kuhn, 2008; Voorhees and Orlowski, 2006). In this regard, BLBC patients may represent a cancer sub-group that might benefit from such a

therapeutic approach. Three of the probe sets comprising the Basal 14 signature bind to transcripts that encode ribosomal protein L3 (RPL3). While it seems likely that this gene is involved in mRNA translation, implying that BLBCs with high levels of protein synthesis are associated with poor patient outcome, the role of RPL3 in cancer is uncharacterized. The genes representing transcripts whose expression was related to good survival are largely uncharacterized in regards to roles in tumor cell biology. Lymphocyte cytosolic protein 1 (LCP1), which is likely expressed by tumor infiltrating lymphocytes, might represent a readout of the extent of tumor lymphocyte infiltrate. This suggests that patient outcome may be influenced by host immune response, where infiltrating immune cells, such as lymphocytes, within a tumor indicate a good prognosis. Indeed, similar observations have been made by multiple other groups in the context of ER negative breast tumors (Teschendorff *et al.*, 2007; Yau *et al.*). Taken together, these data highlight the diverse biology of the genes comprising the Basal 14 signature and provide a scientific rationale for new lines of research aimed at developing BLBC specific therapies.

Several issues remain to be addressed for the Basal 14 signature to be a useful clinical tool. Our conclusions are based on the analysis of retrospective data, which limits its clinical value. Moreover, the validation cohort we used to test the predictive accuracy of the Basal 14 signature was relatively small. Finally, many of the patients in our data-set had incomplete clinical data, making it impossible to learn whether the Basal 14 signature was independently prognostic in the context of other additional factors such as patient age, tumor size, tumor grade, etc. However, it is important to note that previous

reports suggest that factors such as tumor size, tumor grade, extent of vascular invasion, and patient age show little relationship to patient outcome in the context of BLBC especially in lymph-node negative patients(Hudis and Gianni; Rakha *et al.*, 2007). Indeed, the only standard clinical variable that is consistently prognostic in BLBC appears to be nodal status(Hernandez-Aya *et al.*; Rakha *et al.*, 2007). Interestingly, we found that the Genomic Grade Index, a genomic based measurement of tumor grade showed no capacity to stratify BLBC patients into good and poor outcome groups. Subsequent validation of the Basal 14 signature will need to be completed in larger cohorts of patients that include such multivariate analyses. In this regard, a major focus of our research is the optimization of the Basal 14 signature for use on breast tumor tissue that is routinely available after surgery, such as formalin fixed paraffin embedded tumor blocks.

No rigorously validated assay exists to guide prognosis of patients specifically with BLBC. Indeed, the data we present here suggests that the possibility of developing such a test exists. Future experiments will aim to extend these findings in additional retrospective cohorts of patients with BLBCs and ultimately in a prospective based clinical trial aimed at sparing low risk BLBCs patients from detrimental and unnecessary adjuvant chemotherapy.

Methods

We used a four-step approach to complete proof-of-principle experiments to show that gene-expression signatures can be identified and used to classify patients with BLBCs into good and poor outcome groups.

1. We assembled a large cohort of 995 breast tumor gene expression profiles for which clinical follow-up data was available.
2. We classified each tumor on the basis of its 'intrinsic' molecular subtype from which we generated a new dataset consisting of only BLBCs.
3. We used a subset of BLBCs to iteratively identify several prognostic gene signatures, and used cross-validation to identify the optimal signature for patient outcome classification.
4. We validated our optimized signature prospectively on an independent subset of basal breast tumors with accompanying gene expression profiles and clinical follow-up data.

Collecting Microarray Data

We analyzed the gene expression profiles of 5 independent external datasets, obtained using Affymetrix HG-U133A GeneChips arrays, which have been deposited in the Gene Expression Omnibus (GEO); accession numbers GSE1456, GSE2034, GSE3494, GSE6532, and GSE7390. Together these datasets provided expression profiles of 1,077 human breast tumor samples. All gene expression profiles were normalized with frozen Robust Multi-Array Analysis (fRMA), a procedure that allows one to pre-process

microarrays individually or in small batches and to then combine the data into a single comparable dataset for further analyses (McCall *et al.*). To remove batch effect from the combined dataset, we used the ComBat method, which uses an Empirical Bayes method to adjust for potential batch effects in the dataset (Johnson *et al.*, 2007)(http://genepattern.broadinstitute.org), and computed Pearson correlation coefficients for pair-wise comparisons of samples using 68 house-keeping probe sets; only samples exhibiting correlations higher than 0.95 with at least half of the dataset were selected for further classification. The latter filtering method yielded a dataset comprising 995 human breast tumor samples.

Tumor Classification

Each of the selected 995 samples described above, were classified as basal-like, HER2+, Luminal A, Luminal B, claudin-low or normal-like by assigning it to a cluster representing the subtype to which it had the highest Pearson correlation (Herschkowitz *et al.*, 2007; Perou *et al.*, 2000; Sorlie *et al.*, 2001). The correlation was computed using the subset of 1,500 averaged and median-centered 'intrinsic' genes (Parker *et al.*, 2009)common to both our dataset (Affymetrix Human Genome U133A Array) and the dataset used by Parker *et al.* (Stanford Microarray). For robustness, only tumors exhibiting a correlation higher than 0.3 with any of the molecular subtypes were used for further analysis. This led to the classification of 137 breast tumors into the basal-like molecular subtype yielding a group of 134 tumors with useable clinical follow-up data. We randomly separated the 134 patients with basal breast tumors; approximately 2/3

(n=85) were taken for signature training purposes (training set), whereas and the remaining 1/3 (n=49) was used as an independent validation set.

Binary regression

Identification of the prognostic signature was completed using the Bayesian binary regression algorithm BinReg ver2.0. The binary regression software (BinReg2.0) was downloaded from http://web.duke.edu/~dinbarry/BINREG/ and was used as a MATLAB plug-in (West *et al.*, 2001). In most cases, we used disease free survival (DFS) as the relevant clinical variable, however, in some cases only distant metastasis free survival (DMFS) was available within a patient's clinical annotation. In these cases we counted DMFS as DFS. We used 5 years DFS as the clinical endpoint for these studies.

Assessing signature performance

Survival differences between predicted good and poor outcome groups were evaluated with Kaplan-Meier survival curves and a log-rank test for significance. Many standard prognostic clinical variables (node, grad, size, age…, etc) were unavailable in the GEO files associated with the patients used in this study, thus a limitation of this study is that we were not able to test the capacity of the Basal 14 signature to remain prognostic in the context of standard prognostic factors.

Comparison of the Basal 14 signature with other genomic based predictors

We tested multiple additional prognostic signatures on the 49 patient validation cohort: Genomic Grade(Sotiriou *et al.*, 2006), NKI-70(van de Vijver *et al.*, 2002),

Recurrence score(Paik *et al.*, 2004), Wound response(Chang *et al.*, 2005), Triple negative(Yau *et al.*) and MS-14 (Tutt *et al.*, 2008). For cross platform comparisons with other gene signatures, signature elements were mapped by Unigene IDs to Affymetrix HG-U133A GeneChip arrays for testing in the 49 patient validation set. The expression values for each gene were transformed such that the mean was 0 and the standard deviation was 1. A signature index was calculated for each patient as follows:

$$\frac{\sum_{i \in P} x_i}{n_P} - \frac{\sum_{i \in N} x_j}{n_N}$$

Where x is the transformed expression, n is the number of genes that could be mapped to the Affymetrix HG-U133 arrays, P is the set of probes with reported positive correlation to poor outcome, and N is the set of probes with reported positive correlation to good outcome. For each signature, patients were divided into high and low signature index groups using tertiles (Haibe-Kains *et al.*, 2008).

References

(1998). 1997 update of recommendations for the use of tumor markers in breast and colorectal cancer. Adopted on November 7, 1997 by the American Society of Clinical Oncology. *J Clin Oncol* **16:** 793-5.

Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J *et al* Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* **7:** e1000279.

Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM *et al* (2006). Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* **98:** 1183-92.

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T *et al* (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102:** 3738-43.

Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK *et al* (2008). Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res* **14:** 1368-76.

Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA *et al* (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* **13:** 4429-34.

Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G *et al* (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* **14:** 5158-65.

Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B *et al* (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* **13:** 3207-14.

Goulet I, Boisvenue S, Mokas S, Mazroui R, Cote J (2008). TDRD3, a novel Tudor domain-containing protein, localizes to cytoplasmic stress granules. *Hum Mol Genet* **17:** 3055-74.

Haibe-Kains B, Desmedt C, Sotiriou C, Bontempi G (2008). A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* **24:** 2200-8.

Hayes DF, Trock B, Harris AL (1998). Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? *Breast Cancer Res Treat* **52:** 305-19.

Hernandez-Aya LF, Chavez-Macgregor M, Lei X, Meric-Bernstam F, Buchholz TA, Hsu L *et al* Nodal status and clinical outcomes in a large cohort of patients with triple-negative breast cancer. *J Clin Oncol* **29:** 2628-34.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z *et al* (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8:** R76.

Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA *et al* (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24:** 4236-44.

Hudis CA, Gianni L Triple-negative breast cancer: an unmet medical need. *Oncologist* **16 Suppl 1:** 1-11.

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

Karn T, Pusztai L, Holtrich U, Iwamoto T, Shiang CY, Schmidt M *et al* Homogeneous datasets of triple negative breast cancers enable the identification of novel prognostic and predictive signatures. *PLoS One* **6:** e28403.

Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H *et al* (2007). Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* **9:** R65.

McCall MN, Bolstad BM, Irizarry RA Frozen robust multiarray analysis (fRMA). *Biostatistics* **11:** 242-53.

Mulligan AM, Pinnaduwage D, Bull SB, O'Malley FP, Andrulis IL (2008). Prognostic effect of basal-like breast cancers is time dependent: evidence from tissue microarray studies on a lymph node-negative cohort. *Clin Cancer Res* **14:** 4168-74.

Nagahata T, Onda M, Emi M, Nagai H, Tsumagari K, Fujimoto T *et al* (2004). Expression profiling to predict postoperative prognosis for estrogen receptor-negative

breast cancers by analysis of 25,344 genes on a cDNA microarray. *Cancer Sci* **95:** 218-25.

Orlowski RZ, Kuhn DJ (2008). Proteasome inhibitors in cancer therapy: lessons from the first decade. *Clin Cancer Res* **14:** 1649-57.

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M *et al* (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351:** 2817-26.

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T *et al* (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27:** 1160-7.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA *et al* (2000). Molecular portraits of human breast tumours. *Nature* **406:** 747-52.

Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO (2007). Prognostic markers in triple-negative breast cancer. *Cancer* **109:** 25-32.

Ridolfi RL, Rosen PP, Port A, Kinne D, Mike V (1977). Medullary carcinoma of the breast: a clinicopathologic study with 10 year follow-up. *Cancer* **40:** 1365-85.

Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E *et al* (2009). T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res* **11:** R15.

Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L *et al* A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res* **13:** R97.

Sabatier R, Finetti P, Mamessier E, Raynaud S, Cervera N, Lambaudie E *et al* Kinome expression profiling and prognosis of basal breast cancers. *Mol Cancer* **10:** 86.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98:** 10869-74.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100:** 8418-23.

Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J *et al* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* **98:** 262-72.

Staaf J, Ringner M, Vallon-Christersson J, Jonsson G, Bendahl PO, Holm K *et al* Identification of subtypes in human epidermal growth factor receptor 2--positive breast cancer reveals a gene signature prognostic of outcome. *J Clin Oncol* **28:** 1813-20.

Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C (2007). An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol* **8:** R157.

Tutt A, Wang A, Rowland C, Gillett C, Lau K, Chew K *et al* (2008). Risk estimation of distant metastasis in node-negative, estrogen receptor-positive breast cancer patients using an RT-PCR based prognostic expression signature. *BMC Cancer* **8:** 339.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M *et al* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530-6.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

Van Troys M, Huyck L, Leyman S, Dhaese S, Vandekerkhove J, Ampe C (2008). Ins and outs of ADF/cofilin activity and regulation. *Eur J Cell Biol* **87:** 649-67.

Vincent-Salomon A, Gruel N, Lucchesi C, MacGrogan G, Dendale R, Sigal-Zafrani B *et al* (2007). Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. *Breast Cancer Res* **9:** R24.

Voorhees PM, Orlowski RZ (2006). The proteasome and proteasome inhibitors in cancer therapy. *Annu Rev Pharmacol Toxicol* **46:** 189-213.

Wang W, Eddy R, Condeelis J (2007). The cofilin pathway in breast cancer invasion and metastasis. *Nat Rev Cancer* **7:** 429-40.

Wang W, Goswami S, Lapidus K, Wells AL, Wyckoff JB, Sahai E *et al* (2004). Identification and testing of a gene expression signature of invasive carcinoma cells within primary mammary tumors. *Cancer Res* **64:** 8585-94.

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F *et al* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365:** 671-9.

West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R *et al* (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A* **98:** 11462-7.

Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B *et al* (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* **10:** R65.

Xie Y, Wolff DW, Wei T, Wang B, Deng C, Kirui JK *et al* (2009). Breast cancer migration and invasion depend on proteasome degradation of regulator of G-protein signaling 4. *Cancer Res* **69:** 5743-51.

Yau C, Esserman L, Moore DH, Waldman F, Sninsky J, Benz CC A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *Breast Cancer Res* **12:** R85.

Table and Figure legends

Table 1: Features comprising the optimal 14-gene signature

Table 2: Survival characteristics of the 49 patient validation cohort.

Table 3: Survival characteristics of the 26 patient chemo-naïve validation cohort

Figure 1. Human breast tumors cluster into 6 distinct molecular subtypes of breast cancer with differences in patient survival. **A)** Hierarchical clustering of 547 breast tumors using the 'intrinsic' gene set separates tumors into the 6 molecular subtypes of human breast tumors. **B)** Kaplan-Meier survival analysis of patients comprising each of the molecular subtypes

Figure 2. 14 probe sets optimally separate patients into good and poor survival groups. **A)** Experimental strategy to identify an optimal signature to separate patients with BLBC into high and low risk groups. **B)** Comparison of relative risk between leave-one-out cross-validation predicted high and low risk groups for n length signatures (n=1,2,3…,50). 14 probe sets produces maximal risk separation between high and low risk groups (blue arrow).

Figure 3. The Basal 14 signature accurately predicts outcome in independent patients with BLBC. A) Rug plot (distribution of predicted probabilities) of proportion of patients experiencing disease relapse increases linearly with probability of relapse predicted by Basal 14 signature. **B)** ROC curve to assess the accuracy of the Basal 14 signature in the validation cohort (AUC: 0.76, p = 0.003). **C)** Kaplan-Meier survival

analysis with the validation (HR: 4.7, [CI95: 1.8-12.3], p = 0.0017, Log-rank test). **D)** Kaplan-Meier survival analysis with chemotherapy naïve patients (HR: 4.4, [CI95: 1.1-16.7], p = 0.03, Log-rank test).

Figure 4. Other reported prognostic signatures fail to predict patient outcome in the context of BLBC. We calculated a signature index for the **A)** Basal 14, **B)** Genomic Grade Index, **C)** NKI-70, **D)** Recurrence Score, **E)** CSR/Wound response, **F)** Triple Negative and **G)** MS-14 signatures. Only the Basal 14 signature was prognostic in the validation cohort of BLBC patients HR: 4.7 [CI95: 1.8-12.3], p = 0.0017, Log-rank test). Although, the Triple negative signature did trend to significance (HR: 2.0 [CI95: 0.8-5.4], p = 0.15, Log-rank test).

Figure 5. Basal 14 signature is prognostic in the basal and ERBB2 molecular sutbypes of breast cancer. Prognostic capacity of the Basal 14 signature was evaluated in the **A)** luminal A, **B)** luminal B, **C)** claudin low, **D)** Normal, and **F)** ERBB2 molecular subtypes of breast cancer. Notably, the Basal 14 signature was prognostic in patients with the ERBB2 molecular subtype of breast cancer (HR: 2.8 [CI95: 1.3-6.5], p = 0.01, Log-rank test).

Supplementary figure 1. ROC curves for multiple prognostic signatures reported in the literature. ROC curves for **A)** Basal 14, **B)** Genomic Grade Index, **C)** NKI-70, **D)** Recurrence Score, **E)** CSR/Wound response, **F)** Triple Negative and **G)** MS-14 signatures. Only the Basal 14 (AUC: 0.77, p = 0.003) and Triple negative (AUC: 0.7, p = 0.02) signatures produced an AUC that was statistically significantly greater than 0.5.

Table 1: Features comprising the optimal 14-gene signature

Correlation	Affymtetrix Probe	Description
+	201022_s_at	destrin (actin depolymerizing factor), **DSTN**
+	203072_at	myosin IE, **MYO1E**
+	208089_s_at	tudor domain containing 3, **TDRD3**
+	204338_s_at	regulator of G-protein signaling 4, **RGS4**
+	220719_at	hypothetical protein FLJ13769, **FLJ13769**
+	212039_x_at	ribosomal protein L3, **RPL3**
+	211073_x_at	ribosomal protein L3, **RPL3**
+	201217_x_at	ribosomal protein L3, **RPL3**
+	208538_at	acidic (leucine-rich) nuclear phosphoprotein 32 family, member C, **ANP32C**
-	217434_at	melanocortin 2 receptor (adrenocorticotropic hormone), **MC2R**
-	216143_at	MRNA; cDNA DKFZp434L092 (from clone DKFZp434L092), ---
-	221306_at	G protein-coupled receptor 27, **GPR27**
-	204544_at	Hermansky-Pudlak syndrome 5, **HPS5**
-	208885_at	lymphocyte cytosolic protein 1 (L-plastin), **LCP1**

Table 2: Survival characteristics of the 49 patient validation cohort.

Validation cohort (n=49)			
Risk Category	**# Patients**	**% Patients**	**% Disease free survival (5 yr)**
Low	16	33	94
High	33	67	48

Table 3: Survival characteristics of the 26 patient chemo-naïve validation cohort

Chemo-naïve validation cohort (n=26)			
Risk Category	**# Patients**	**% Patients**	**% Disease free survival (5 yr)**
Low	6	23	100
High	20	77	50

Figure 1. Hallett, R. *et al.*

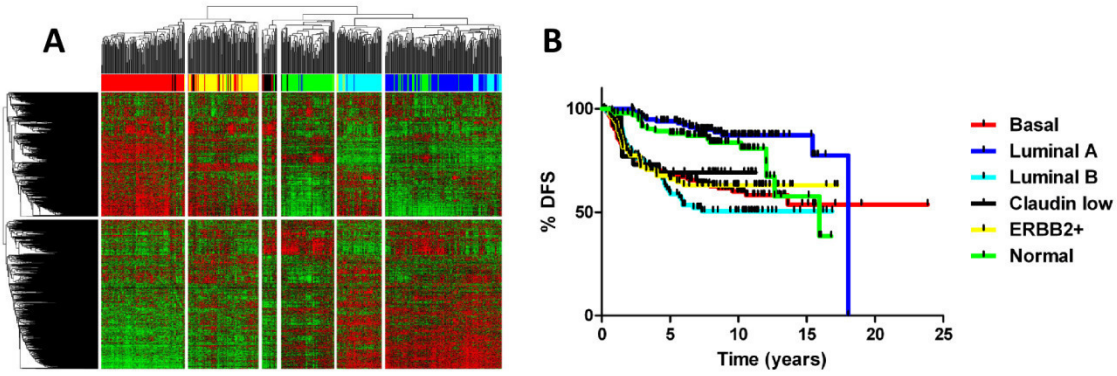


Figure 2. Hallett, R. *et al.*

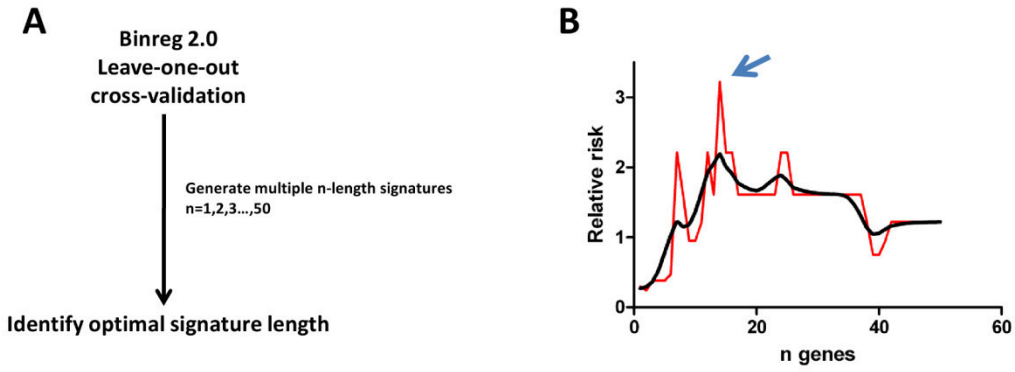


Figure 3. Hallett, R. *et al.*

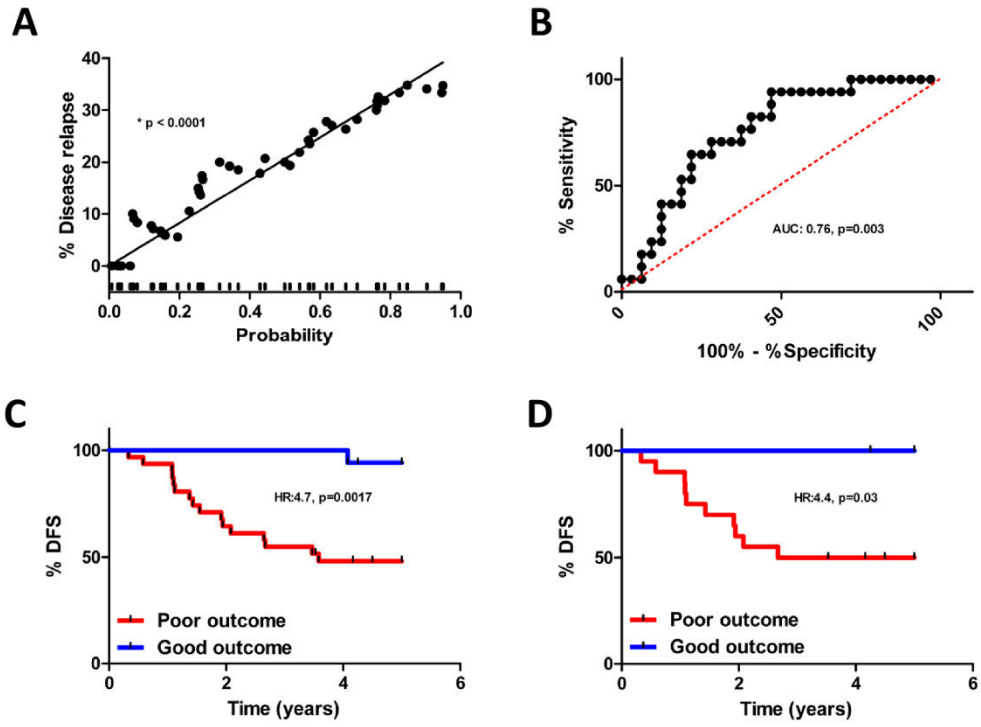


Figure 4. Hallett, R. *et al.*

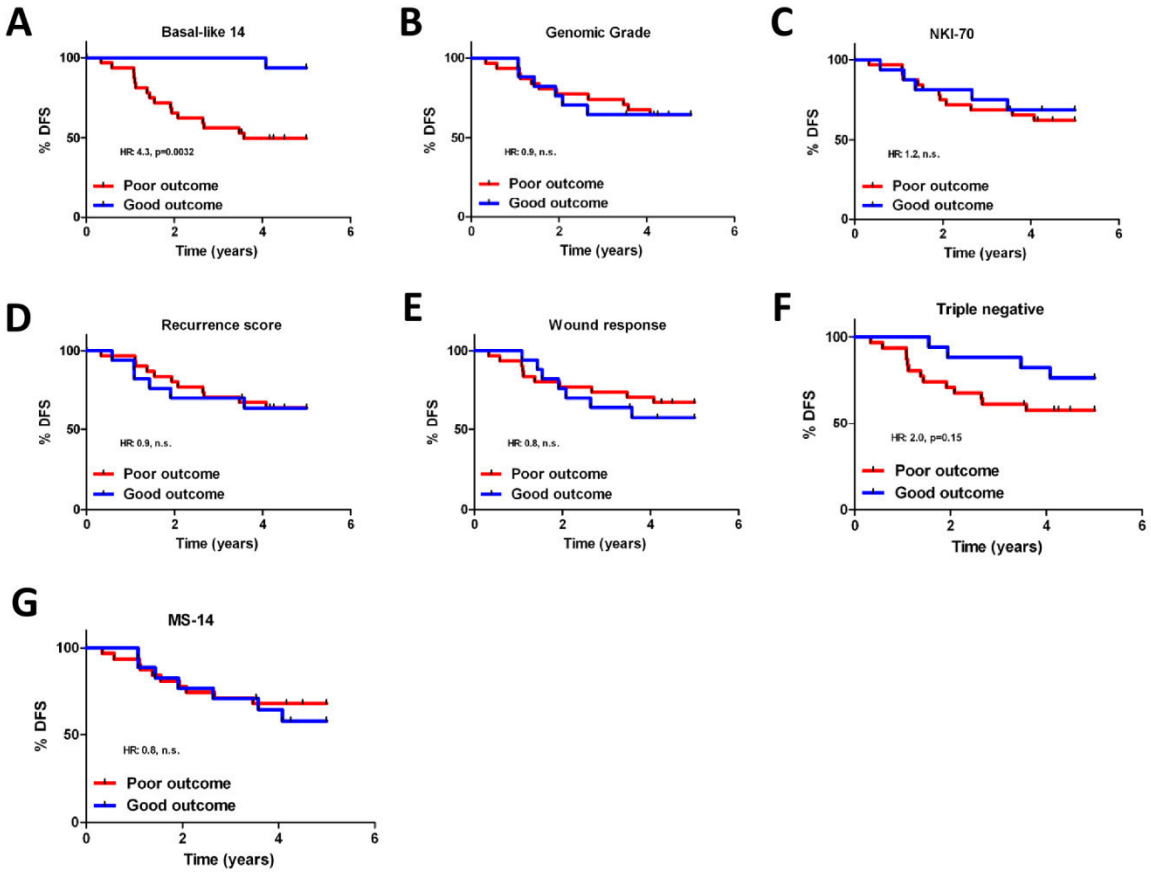
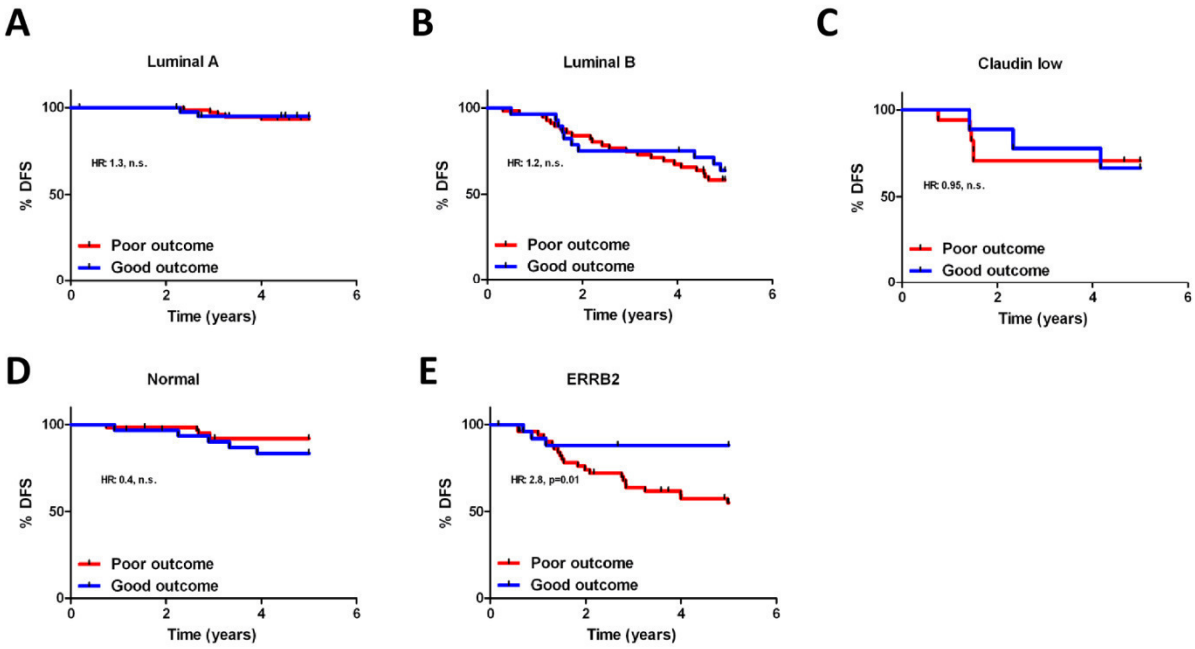
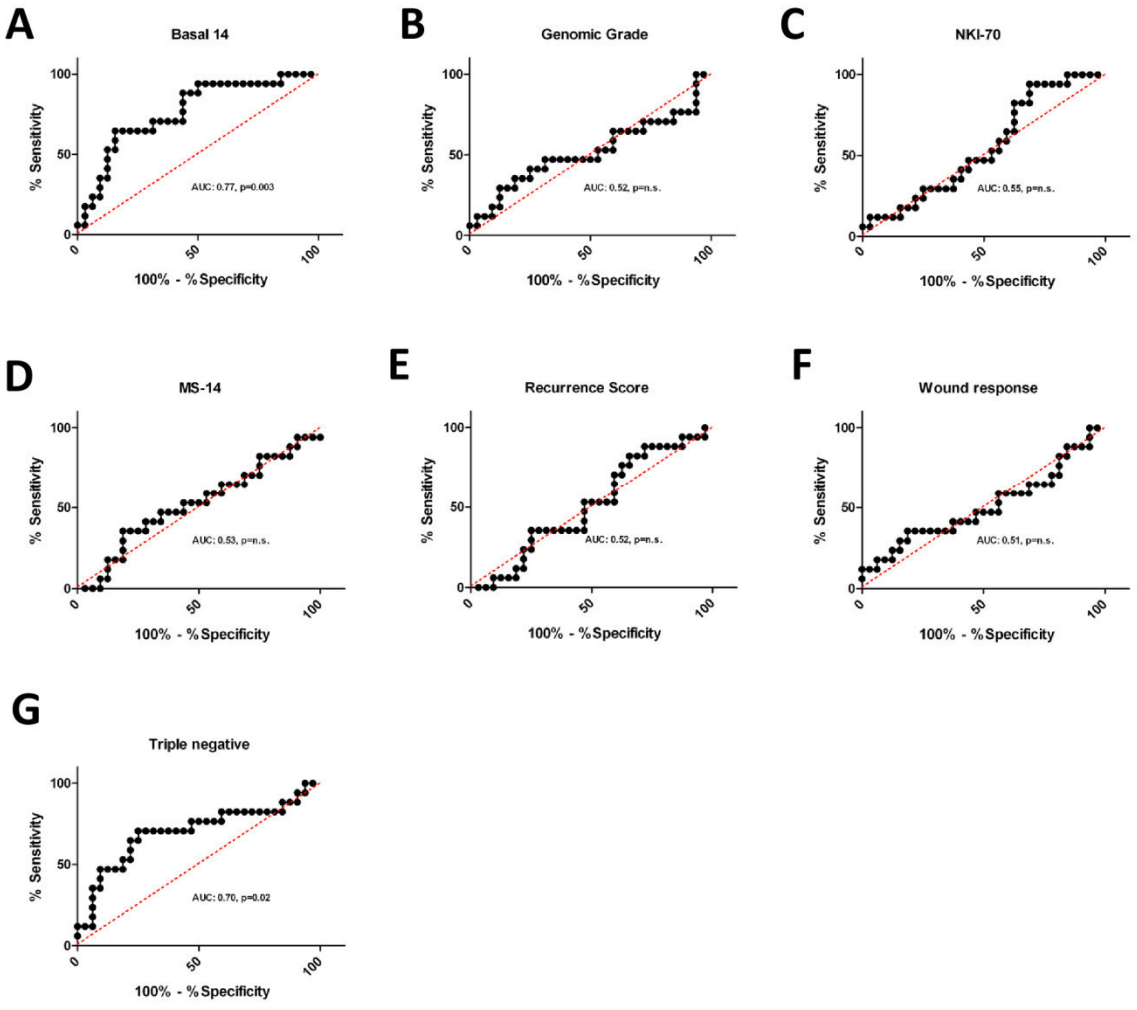


Figure 5. Hallett, R. *et al.*



Supplementary figure 1. Hallett, R. *et al.*



CHAPTER 3

Introduction

There is significant molecular and cellular diversity among human breast tumors. Indeed, this heterogeneity is evident from histopathololologic features and differences in ER/PR/HER2 status as well as more recent classification schemes based on the expression of large numbers of genes (Perou *et al.*, 2000; Sorlie *et al.*, 2001; Sorlie *et al.*, 2003; Sotiriou and Pusztai, 2009). Importantly, these data indicate that breast cancer is an imprecise definition that embodies many molecularly distinct neoplastic disorders that share a common normal breast tissue origin.

The capacity to more accurately define breast cancers and identify tumor subgroups that represent more homogeneous disease entities, provides a framework to increase our understanding of these diseases and provides opportunities to focus treatment options for patients. To this end, investigators have completed relatively large gene expression studies and identified patterns in gene expression that reproducibly stratify breast tumors into each of 5 molecular groups. These groups, known as the basal-like, ERBB2+, normal-like, luminal A and luminal B molecular subtypes of breast cancer, were originally described by Perou *et al.(Perou et al., 2000)* in 2000, and define disease groups with distinct clinical behaviors, thus providing a basis for improved taxonomy of breast cancer (Blows *et al.*; Desmedt *et al.*, 2008; Herschkowitz *et al.*, 2007; Parker *et al.*, 2009; Rouzier *et al.*, 2005; Smid *et al.*, 2008; Sorlie *et al.*, 2001; Sorlie *et al.*, 2003; Sotiriou and Pusztai, 2009; Wirapati *et al.*, 2008). It is possible that further refinement of the 'intrinsic' classification by Perou *et al.,* would delineate additional

molecular classes of breast cancer, and provide additional clinical value beyond traditional techniques. For example, ER+ tumors generally fall into the luminal A and B molecular subtypes, characterized by expression of the ER as well as cytokeratins typically expressed by luminal epithelial cells (Perou *et al.*, 2000; Sorlie *et al.*, 2001). However, more recent studies suggest that as many 12 molecular subgroups of ER+ breast cancer exist, demonstrating that the luminal A and B stratification of ER+ breast tumors does not fully capture the biological complexity of these tumors {Gatza, 2010 #284}. Indeed, further dissection of ER+ breast tumors into additional relevant disease subgroups would likely provide further insight into the mechanisms that underlie these tumors, as well as prevent carefully planned studies being confounded by the heterogeneity found among un-grouped or sub-optimally grouped populations of ER+ breast tumors. Hence, the identification and characterization of additional subgroups of ER+ breast tumors would focus treatment options for patients with ER+ breast tumors, because therapy could be rationally applied based on specific molecular characteristics of the patient's tumor.

We hypothesized that the biology of ER+ tumors comprised both estrogen-dependent and -independent components, and furthermore, that investigation and characterization of the estrogen independent component might provide a means to stratify ER+ tumors into different distinct disease subgroups. To this end, we used publicly available data to identify 'estrogen independent' genes in ER+ breast tumors and subsequently identified subgroups of ER+ tumors based on molecular differences between tumors identified by these genes. Importantly, we reproducibly identified 8

subgroups of ER+ breast tumors that exhibited distinct clinical behavior as well as biology. Moreover, we show that these subgroups have specific responses to therapeutic compounds *in vitro,* as well as in human breast cancer patients. Taken together, these data support the existence of 8 distinct subgroups of ER+ breast cancer, and advance efforts to increase the precision of therapeutic intervention in human breast cancer patients.

Manuscript #2

Title: Estrogen independent gene expression defines clinically relevant subgroups of estrogen receptor positive breast cancer

Authors: Robin M. Hallett[1], and John A. Hassell[1]*

1. Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5

*Correspondence and requests for materials should be addressed to J.A.H. (hassell@mcmaster.ca)

Acknowledgements

This work was generously supported by grants from the Canadian Stem Cell Network, the Ontario Institute for Cancer Research and the Canadian Breast Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We wish to acknowledgements helpful discussion from Drs. Greg Pond and Anita Bane throughout the course of this work.

Author contributions

Robin Hallett – Conceived, planned, analyzed and performed most of the experiments in the paper, wrote the manuscript

John Hassell – Helped write manuscript, provided critical feedback for project

Abstract

Human breast cancer represents a significantly heterogeneous disease. Indeed, the latter is readily apparent based on the variable expression of the Estrogen Receptor (ER), Progesterone Receptor (PR) and expression of HER2 in patient tumors. Moreover, global gene expression profiling measurements have been used to classify tumors into multiple molecular subtypes. The capacity to define subtypes of breast tumors provides a framework to enable improved understanding of the mechanisms of breast oncogenesis, as well as to provide opportunities for improved targeted therapeutic intervention in patients. We hypothesized that the biology of ER+ tumors comprised both estrogen-dependent and -independent components, and that investigation of the estrogen-independent component might provide a means to stratify ER+ tumors into additional distinct subgroups. We used publicly available gene expression profiling data to identify 'estrogen independent' genes in ER+ breast tumors. Importantly, based on the expression of these genes, we identified 8 subgroups of ER+ breast tumors that exhibit distinct clinical behaviors and biology. Moreover, we show that breast tumor cell lines representative of these ER+ subgroups have subtype specific response to therapeutic compounds *in vitro,* and that this also applies to breast cancer patients. Taken together, these data support the existence of 8 distinct subgroups of ER+ breast cancer and suggest that knowledge of the ER+ subgroup status of patient samples have the potential to guide therapy choice.

Results

Identification of estrogen independent genes and distinct subgroups of ER+ breast cancer

The goal of this study was to enable classification of ER+ breast tumors on the basis of genes whose expression is related to the estrogen independent biology of ER+ tumors. To this end, we took advantage of the gene expression profiles of 58 ER+ breast tumors biopsied from post-menopausal women before and after treatment with letrozole (n=116, [20]). Because letrozole treatment induces estrogen deprivation in tumors of post-menopausal women, genes whose expression showed minimal variation after letrozole treatment can be considered estrogen independent. To identify estrogen-independent genes, we calculated between/within (b/w) scores for each probe set, which were measurements of probe set variation observed between different primary tumors relative to the variation observed within paired samples pre- and post-treatment. In this fashion, probe sets with high b/w scores showed greater variation between different tumors than between treatment paired tumor samples, whereas genes with low b/w scores showed greater variation within treatment paired tumors samples than between different tumors. Therefore, probe sets with high b/w scores were likely not influenced by estrogen deprivation and also showed variable expression among the different tumors prior to treatments, suggesting that they are related to differences in the estrogen independent biologies of ER+ tumors (Fig. 1A). We selected the top 5,000 estrogen independent probe (highest b/w scores) sets for further analysis (Fig.1B).

To investigate whether the expression of the estrogen independent probe sets could capture the phenotypic complexity of ER+ breast tumors we completed unsupervised clustering using non-negative matrix factorization (NMF) [21]. NMF is an efficient method to identify molecular patterns that is readily applicable to gene expression data, and therefore can be used as a powerful means for class discovery. In short, NMF identifies metagenes, or distinct gene expression patterns, which are used to determine the optimal value for k, where k represents the number of sample subgroup clusters by calculating a cophenetic co-efficient for each value of k. In short, we applied NMF (for k=2-10) to gene expression data representing 327 primary ER+ breast tumors (GSE6532, 133A arrays, [22] filtered such that only the 5000 estrogen independent probe sets were used for class identification. This data set optimally fell into 8 clusters, designated groups 1–8 (Fig. 1C). To identify marker genes for each subgroup, we used prediction analysis of microarrays (PAM, [23]), to identify the top 50 marker genes for each group (Fig. 1D, Supplementary table 2). To learn whether these groups might represent different disease classes, we compared distant metastasis free survival (DMFS) among the various groups. Interestingly, some subgroups displayed excellent long term outcomes, whereas other groups did not (Fig. 1E, *p=0.03, Log-rank test). For example, 10 year DMFS in group 7 patients was 90%, whereas in group 4 patients it was only 52%. Importantly, these patients were uniformly chemotherapy naïve, suggesting that these differences in survival are likely related to the natural progression of their disease, rather than response to chemotherapy.

To learn whether different signaling pathways might be associated with each group we tested whether subgroup genes were over-represented within interaction partners that comprise distinct pathways in the NCI-Nature pathway interaction (Table 1) [24]. Interestingly, group 1&4 genes were both enriched in genes that comprise the EGFR signaling pathway, a known marker of poor outcome in breast tumors [25], providing a potential biological link to the poor DMFS observed in these groups. By contrast, group 7 genes were enriched in immune-signaling pathways, such as CXCR4-mediated signaling and downstream signaling in CD8 T-cells, suggesting that group 7 tumors comprised a relatively high percentage of infiltrating immune cells, possibly accounting for their excellent survival characteristics. Interestingly, many additional therapeutically actionable signaling pathways were over-represented among other subgroups, suggesting that tumor subgroup membership could inform patient treatment. Taken together, these data suggest that ER+ breast tumors can be grouped into 8 distinct disease subgroups with significant differences in clinical outcome, as well as with unique biology and signalling.

A framework for ER+ breast tumor classification

To confirm that the previously observed subgroups were indeed representative of ER+ breast tumors, we investigated their prevalence in an independent validation dataset comprising 805 ER+ chemotherapy naïve, breast cancer patients. We applied the 400 probe set classifier (top 50 probe sets of each subgroup, described above) to the validation set, to separate these tumors into the 8 subgroups identified above (Fig 2A).

Based on the expression of the 400 probe sets, each sample was assigned into one of the 8 groups, using PAM. Notably, greater than 80% (n=653) of the tumors in the validation set were assigned with a probability higher than 80% of belonging to one of the 8 subgroups, suggesting this classification framework is reasonably robust. The DMFS characteristics of patients comprising the various 8 groups were found to be highly coincident between the original (n=327) and validation (n=805) cohorts (Fig. 2B, supplementary figure 1). For instance, 10 year DMFS was highest in group 7 and lowest in group 4 for both the original and validation datasets. We also investigated DMFS in patients that comprised the validation groups who were also node negative (Fig. 2C), or were node negative as well as treated with tamoxifen (Fig. 2D). In the latter analysis, it appeared that patients whose tumors belonged to groups 3, 6, and 7 experienced excellent DMFS (92%), whereas patients whose tumors belonged to groups 1, 2, 4, 5, and 8 experienced much poorer DMFS (78%, Fig. 2D). The latter patients (ER+, tamoxifen treated, node negative) represent an important patient population as they are eligible for gene expression based tests, such as Oncotype DX [4,26], which are designed to guide of the use of chemotherapy in these patients. These data suggest that the subgroups of ER+ tumors identified here could be used in a similar fashion to identify patients that are unlikely to require additional therapeutic intervention and those that require more aggressive treatment.

Importantly, the 400 probe set classifier robustly classified validation set patients into the subgroups identified in our original dataset. Moreover, the survival characteristics between patients whose tumors were classified into the various subgroups

were highly coincident between the validation and original datasets. Taken together, these data are highly suggestive that the identified subgroups are both clinically relevant and reproducible.

ER+ subgroups enable predictive modeling of anti-cancer drug sensitivity

As described above, the established framework allows classification of ER+ breast tumors into 1 of 8 various subgroups based on patterns in gene expression. Importantly, this framework can be extended to classify ER+ breast tumor cell lines into the same subgroups. We accessed previously described breast tumor cell line gene expression datasets [18,27] and classified ER+ breast tumor cell lines into the 8 subgroups. Among 25 ER+ breast tumor cell lines, 6 of the 8 subgroups were represented by at least 2 cell lines, thus providing experimental models for these subgroups (Fig. 3A). To learn whether the cell lines comprising the various subgroups could model important clinical characteristics, such as response to anti-cancer therapies, we compared the breast tumor cell line sensitivities [18] with that of breast cancer patient response to neoadjuvant chemotherapy data for anthracycline and taxane drugs, using publicly available data [28]. Interestingly, we observed that the cell lines comprising subgroups 1 and 3 displayed divergent sensitivities to anthracyclines and taxanes. For example, the correlation between the $-\log_{10}GI_{50}$s and probability of assignment into subgroup 1 were relatively high for doxorubicin and epirubicin, as well as paclitaxel and docetaxel (Fig. 3B). In contrast, the same correlation was negative for these same agents

in subgroup 3 (Fig. 3B). In fact, subgroup 1 cell lines were either significantly (paclitaxel, *$p<0.05$) or nominally more sensitive to these agents than subgroup 3 cell lines (docetaxel, doxorubicin, epirubicin) (Fig. 3C). We also compared the efficacy of anthracycline and taxane chemotherapy in subgroup 1 and subgroup 3 breast cancer patients. Measuring the efficacy of chemotherapy in a tumor can be accomplished by assessment of response following neoadjuvant treatment with a given treatment regimen [29]. In this fashion, patients with little to no invasive or metastatic disease detected following treatment are classified as having minimal residual cancer burden (RCB0/I), whereas patient's whose tumors show less response and care classified as having greater residual cancer burden (RCBII/III). Importantly, neoadjuvant chemotherapy has been found to be as efficacious as chemotherapy given in the adjuvant setting, and patients who achieve complete pathological response after neoadjuvant intervention generally have an excellent probability of experiencing long-term survival [30,31,32]. Taken together, these data suggest that response to neoadjuvant chemotherapy (pCR/RD) is a relevant clinical model to develop and validate gene signature based predictors of breast tumor response to chemotherapy. To this end, we generated a dataset comprising 248 patients with ER+ breast tumors that also received neoadjuvant anthracycline and taxane chemotherapy. Patient's with low residual cancer burden (RCB0/RCBI) were considered as responders, whereas patients with high residual cancer burden (RCBII/RCBIII) were classified as non-responders [28,29,33]. We observed that subgroup 1 tumors were generally more responsive than subgroup 3 tumors (Fig. 3D, RCB0/I:RCBII/III, Subgroup 1: 8:18, Subgroup 3: 7:90, OR: 3.5, *$p=0.031$). Taken together, these data

reveal that cell lines provide tractable experimental models to study subgroup response to anticancer agents. Indeed, the observation that cell line subgroup sensitivity/resistance to anthracyclines/taxanes was phenocopied in patient subgroup response to neoadjuvant anthracycline/taxane therapy, provides proof-of-principle that cell lines can be used to model subgroup specific responses to both approved and experimental breast cancer therapeutics.

Cell lines belonging to ER+ subgroups show heterogeneous response to clinical and experimental anti-cancer compounds

To extend the findings described above, we assessed the relationship between ER+ breast tumor cell lines comprising our described subgroups and response to additional clinical and experimental anti-cancer compounds. As reported above, subgroups 1-6 were represented by at least 2 cell lines, thus only these subgroups were included in these analyses. Interestingly, many compounds showed subgroup specificity, suggesting that these efforts may inform effective clinical use of these compounds (Fig. 4A). For example, CPT-11, also commonly known as irinotecan, was most effective in cell lines belonging to subgroup 1 relative to cell lines belonging to other subgroups (Fig. 4B, * $p = 0.0091$), and etoposide was most effective in subgroup 1 and 5 cell lines relative to the those comprising subgroups 2,3,4, and 6 (Fig. 4C, *$p=0.0023$). Importantly, these observed patterns could likely be implemented into clinical practice (Fig. 4D), were upon clinical presentation of ER+ breast cancer, a biopsy is used for gene

expression profiling analysis to make a subgroup assignment. Low risk patients, such as node negative subgroup 3, 6, and 7 disease would only receive endocrine therapy, whereas high risk patients, such as those with node positive disease, or those with tumors belonging to subgroups 1,2,4,5, and 8 would receive endocrine therapy as well as additional therapy informed by the subgroup of the tumor. Importantly, many of the agents with subgroup specificity represent standard of care agents, suggesting that strategy described above could be easily applied and tested in clinical settings [34].

Discussion

The capacity to accurately classify heterogeneous populations of breast tumors into more homogeneous disease groups is important for multiple reasons. First, different classes of breast tumors are likely initiated and propagated by distinct disease mechanisms, implying that the study of such mechanisms would be most tractable within the subgroups they underlie. Second, effective treatment of breast cancer patients will hinge on the capacity to accurately define the characteristics of an individual's tumor in order to tailor subsequent treatments to these characteristics. Hence, the precise classification of breast tumors into therapeutically relevant subgroups provides a means to treat individuals with subgroup specific therapies comprising agents to which the tumor is most likely to respond.

In breast cancer, tumors are typically classified based on clinical parameters such as size and grade, as well as the expression of a limited number of histochemical markers, such as the expression of ER, PR and HER2. Albeit limited in number, these markers have important therapeutic implications; expression of the ER predicts benefit from endocrine therapies, whereas overexpression of HER2 predicts benefit from trastuzumab [35,36,37]. More recent experiments have identified 5 major molecular subtypes of breast tumors, which are based on the expression of large numbers of genes [1,2,3]. Importantly, these subtypes have been linked to important clinical characteristics, such as overall survival and response to chemotherapy [3,8]. However, it is unclear the extent to which these molecular subtypes improves classification of breast cancer beyond

traditional methods [3,4,8,38,39]. Hence, additional classification schemes are required to gain additional understanding of the various biological programs that drive breast oncogenesis, as well as increase the efficacy of treatment. Here we suggest that patterns in estrogen independent gene expression likely provides such a scheme, and link such patterns to unique subgroups of ER+ breast tumors with distinct clinical characteristics and sensitivities to anti-cancer therapeutics.

Here, we used the gene expression profiles of 58 matched untreated/treated ER+ breast tumor pairs for estrogen independent gene discovery. The expression of these genes were subsequently used to identify molecular classes of ER+ breast tumors in a discovery cohort of 327 breast tumors. This analysis identified 8 molecular subgroups that comprised patients with distinct outcomes. Importantly, in a validation cohort of 805 ER+ breast tumors more than 80% tumors were classified into the 8 subgroups with high probability (>80%). Moreover, subgroup specific patient outcomes were highly coincident between the discovery and validation cohorts. Taken together, these observations provide strong evidence for the existence of the 8 subgroups. We also extended these findings by using cell line data to identify subgroups likely to respond to therapies. As proof-of-principle, we linked cell line subgroup response to anthracycline and taxane drugs with patient subgroup response to neoadjuvant anthracycline and taxane therapy. These data highlight the possibility of using cell line subgroup membership in conjunction with cell line response data to guide conventional chemotherapy in ER+ breast tumors. Indeed, among other compounds analysed, many showed subgroup specificity suggesting that we uncovered additional subgroup specific responses to

therapeutic compounds. Along these same lines, many therapeutically actionable pathways were over-represented within subgroup specific genes that were used to define the 8 subgroups. These include PLK1 signaling in subgroup 2 tumors [40], FGF and IL6 signaling in subgroup 3 tumors [41,42], telomerase in subgroup 5 tumors [43], and HDAC signaling in subgroup 8 tumors [44,45], among others. From a clinical standpoint, our data suggests that patients with low risk subgroup tumors (subgroups 3,6,7) experience excellent long term survival without adjuvant chemotherapy, and therefore should only require endocrine therapy. Conversely, high risk subgroup patients (subgroups 1,2,4,5,8) could be treated with additional therapy selected based on subgroup membership. That said, significant additional work remains to be completed before these 8 subgroups could be used to inform clinical decision making. This work would include developing a robust and reliable clinical multi-gene assay to optimally assign subgroup membership. Moreover, the predictive power of subgroup guided decision making would need to be validated in prospective clinical trials.

Importantly, the subgroups reported here are defined based on the expression of genes that showed minimal variation in patient tumors that were treated with an aromatase inhibitor [20]. Aromatase inhibitors decrease patient estrogen levels in post-menopausal women by inhibiting aromatase, thereby reducing the synthesis of estrogen from androgenic substrates [46]. In this fashion, treatment with aromatase inhibitors induces an estrogen deprived state within a patient's tumor. Thus, the expression of genes that show minimal variation likely represents estrogen independent biology present within ER+ tumors. Arguably, since ER+ breast cancer patients are treated with

endocrine therapies that inhibit estrogen dependent gene expression, endocrine therapy resistance is driven at least in part, by the biology mediated by estrogen independent genes. The study of endocrine therapy resistance is particularly important, given that approximately 30% of tamoxifen-treated ER+ breast cancer patients suffer relapse within 15 years [35]. To this end, the pathway analysis of subgroup specific genes might lend insight into means of overcoming such resistance. Notably, subgroup 1 and 4 classifier genes were found to be enriched in components of the EGFR signaling pathway. Moreoever, tamoxifen treated subgroup 1 and 4 patients experienced poor survival suggesting that tamoxifen fails to provide durable cure in these subgroups. An interpretation of these data is that EGFR pathway activity provides an escape pathway for ER+ breast tumors from tamoxifen therapy, consistent with previous reports linking EGFR signaling to endocrine therapy resistance [47,48,49].

Ultimately, we propose that the 8 subgroups described here provide a strategy for improved understanding and treatment of ER+ breast tumor biology. We anticipate these data will provide a framework to both guide optimal use of existing therapeutics, as well as gain insight into biological processes that represent relevant targets for development of the next generation of experimental therapies.

Methods

Human breast tumor data sets.

All tumor samples were downloaded from the gene expression omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). This included Letrozole treated tumor samples (GSE5462) [20], the discovery cohort (GSE6532, 133A array samples, n=327 [50], the validation cohort (GSE6532 133 Plus 2.0 array samples n= 87 [50], GSE9195 n=77 [51], GSE17705 n=298 [52], GSE2034 n=209 [53], GSE7390 n=134 [54]). Cell line expression profiles were downloaded from ArrayExpress (E-TABM-157) [27]. Raw data files reprensenting the tumor samples were normalised using using RMA (http://genepattern.broadinstitute.org) [55].

Definition of estrogen independent genes

We calculated within (w, treatment pairs) and between (b) variation (independent primary tumor samples) for all tumors. In this fashion, probe-sets with greater variation in expression between tumors than between treatment paired samples received high b/w scores, and vice versa.

Class discovery

Non-negative matrix factorization was carried out as previously described on the gene pattern website [21]. Prediction analysis of microarrays (PAM) was carried out as described [23], to discover subgroup specific genes (discovery cohort) and to classify samples (validation cohort, cell lines).

Pathway analysis

Pathway analysis was carried out using NCI-Nature pathway interaction database tools (http://pid.nci.nih.gov/index.shtml). The pathway interaction database is a curated database of biomolecular interactions and cellular processes assembled into signaling pathways [24].

Compound data

Cell line sensitivity to compound data was downloaded from (http://www.pnas.org/content/early/2011/10/13/1018854108.abstract) [18]. To learn whether compounds had subgroup specific activity we calculated correlation (Pearson) between the probability of belonging to a given subgroup and the –log10 of the GI50 of chemotherapies and experimental compounds.

Statistical Analysis

Survival analysis and Log-rank tests were used to evaluate survival differences between patient subgroups. Whenever possible, DMFS was used as the clinical endpoint, however in a minority of cases only disease free survival was available. Odds ratios and the Fisher's exact test were used to compare the frequency of responders and non-responders to neoadjuvant chemotherapy. T-tests were used to compare the GI50s of various compounds between cell lines representing the subgroups. Tests were two-sided and a p-value of 0.05 or less was considered statistically significant.

References

Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR *et al* (2006). Basal-like breast carcinomas: clinical outcome and response to chemotherapy. *J Clin Pathol* 59: 729-35.

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S *et al* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603-7.

Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J *et al* Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* 7: e1000279.

Callagy G, Cattaneo E, Daigo Y, Happerfield L, Bobrow LG, Pharoah PD *et al* (2003). Molecular classification of breast carcinomas using tissue microarrays. *Diagn Mol Pathol* 12: 27-34.

Colleoni M, Cole BF, Viale G, Regan MM, Price KN, Maiorano E *et al* Classical cyclophosphamide, methotrexate, and fluorouracil chemotherapy is more effective in triple-negative, node-negative breast cancer: results from two randomized trials of adjuvant chemoendocrine therapy for node-negative breast cancer. *J Clin Oncol* 28: 2966-73.

Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G *et al* (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 14: 5158-65.

Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW *et al* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483: 570-5.

Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD *et al* A pathway-based classification of human breast cancer. *Proc Natl Acad Sci U S A* 107: 6994-9.

Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S *et al* Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci U S A* 109: 2724-9.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z *et al* (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8: R76.

Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z *et al* (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 10: 5367-74.

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T *et al* (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27: 1160-7.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA *et al* (2000). Molecular portraits of human breast tumours. *Nature* 406: 747-52.

Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K *et al* (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 11: 5678-85.

Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG *et al* (2008). Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 68: 3108-14.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98: 10869-74.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100: 8418-23.

Sotiriou C, Pusztai L (2009). Gene-expression signatures in breast cancer. *N Engl J Med* 360: 790-800.

Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B *et al* (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* 10: R65.

Table and Figure legends

Table 1: Enrichment of subgroup genes in within the NCI-Nature curated pathway database

Figure 1. Discovery of estrogen independent genes. A) Experimental strategy to identify estrogen independent genes. **B)** Between/within scores for all probes. **C)** NMF consensus analysis of discovery cohort identifies 8 subgroups of ER+ breast tumors. **D)** Heatmap displaying the top 50 subgroup genes for each subgroup identified by PAM. **E)** DMFS of patients comprising the various subgroups (*p=0.03, Log-rank test).

Figure 2. 8 subgroups can be reproducibly identified in an 805 ER+ breast tumor patient validation cohort. A) Subgroup assignment by PAM of each patient in the validation cohort **B)** Survival characteristics of patients comprising the various subgroups in the validation cohort (*p=0.01, Log-rank test, n=805). **C)** Subgroup survival characteristics of node negative patients within the validation cohort patients (*p=0.02, Log-rank test, n=588). **D)** Subgroup survival characteristics of tamoxifen treated and node negative patients within the validation cohort (*p=0.04, Log-rank test, n=245). **E)** Pooled survival characteristics of patients representing either low risk or high risk disease (HR: 2.55, *p=0.0037, Log-rank test)

Figure 3. ER+ subgourps enable predictive modeling of anti-cancer drug sensitivity. A) Subgroup assignment for 25 ER+ breast tumor cell lines. **B)** Relationship between probability of subgroup membership and sensitivity to anthracycline and taxane drugs for subgroups 1 and 3. **C)** Subgroup 1 cell lines are either statistically (paclitaxel: *p=0.036)

or nominally more sensitive to anthracyclines and taxane drugs (docetaxel: p=0.25, doxorubicin: p=0.071, epirubicin: p=0.1). **D)** Subgroup 1 tumors are more responsive to anthracycline and taxane chemotherapy than subgroup 3 tumors in human breast cancer patients (Odds Ratio: 3.5 [1.2-10.2], *p=0.031, Fisher's exact test).

Figure 4. Subgroup specific response to anti-cancer compounds. A) Heatmap depicting cell line response to anti-cancer agents. **B)** Subgroup 1 cell lines are markedly more sensitive to the agent CPT-11 than the cell lines belonging to the other subgroups (*p=0.0091). **C)** Subgroup 1 and 5 cell lines are markedly more sensitive to etoposide than cell lines comprising the remaining subgroups (*p=0.0023). **D)** Strategy for implementation of the 8 subgroups into management of patients with ER+ breast tumors.

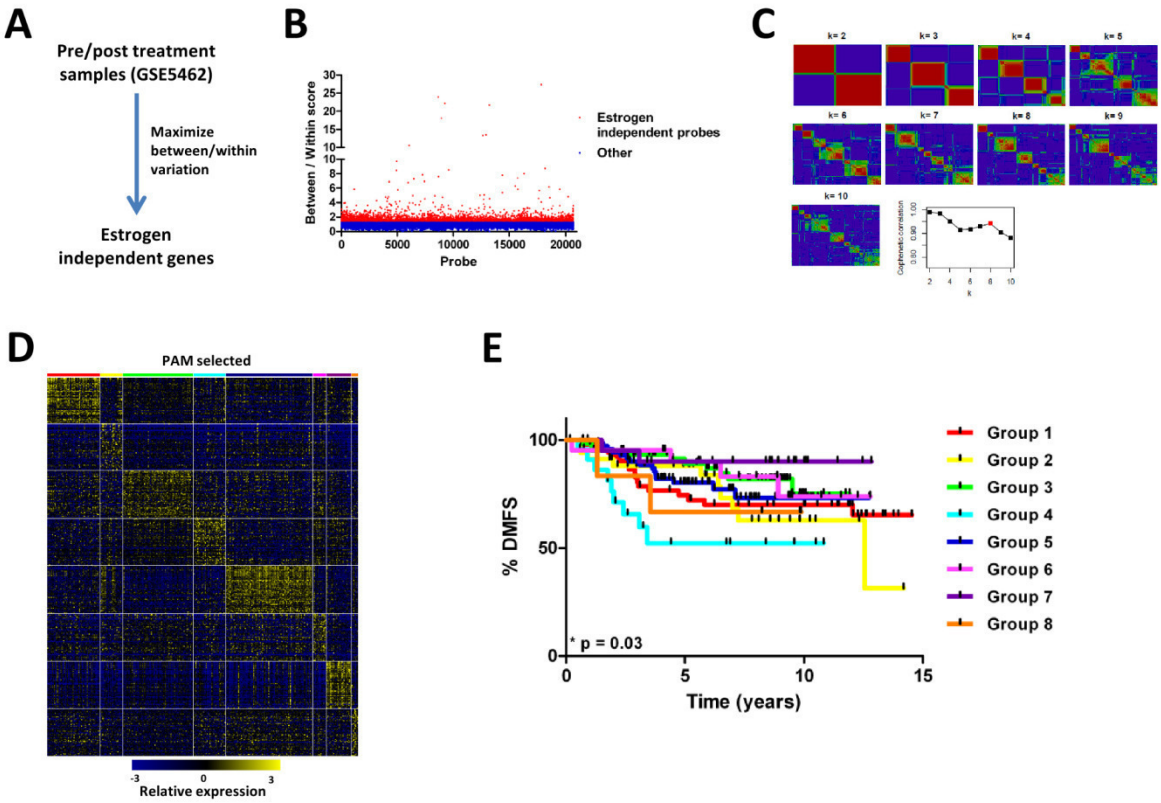
Supplementary Figure 1. Comparison of 10 year-survival between validation and training cohorts.

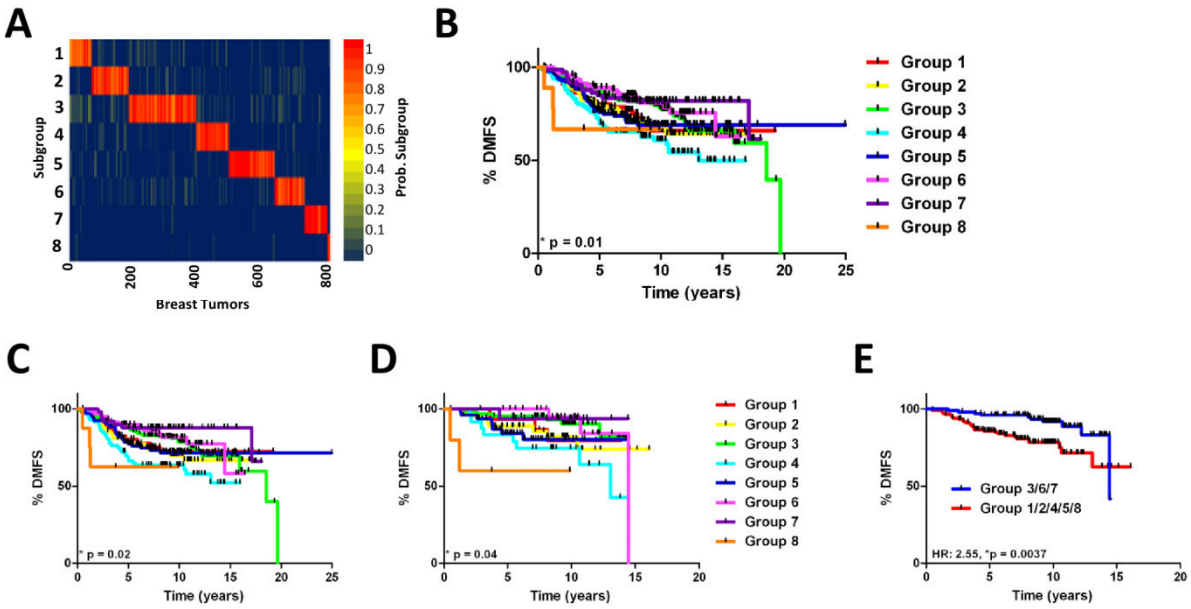
Table 1: Enrichment of subgroup genes in within the NCI-Nature curated pathway database

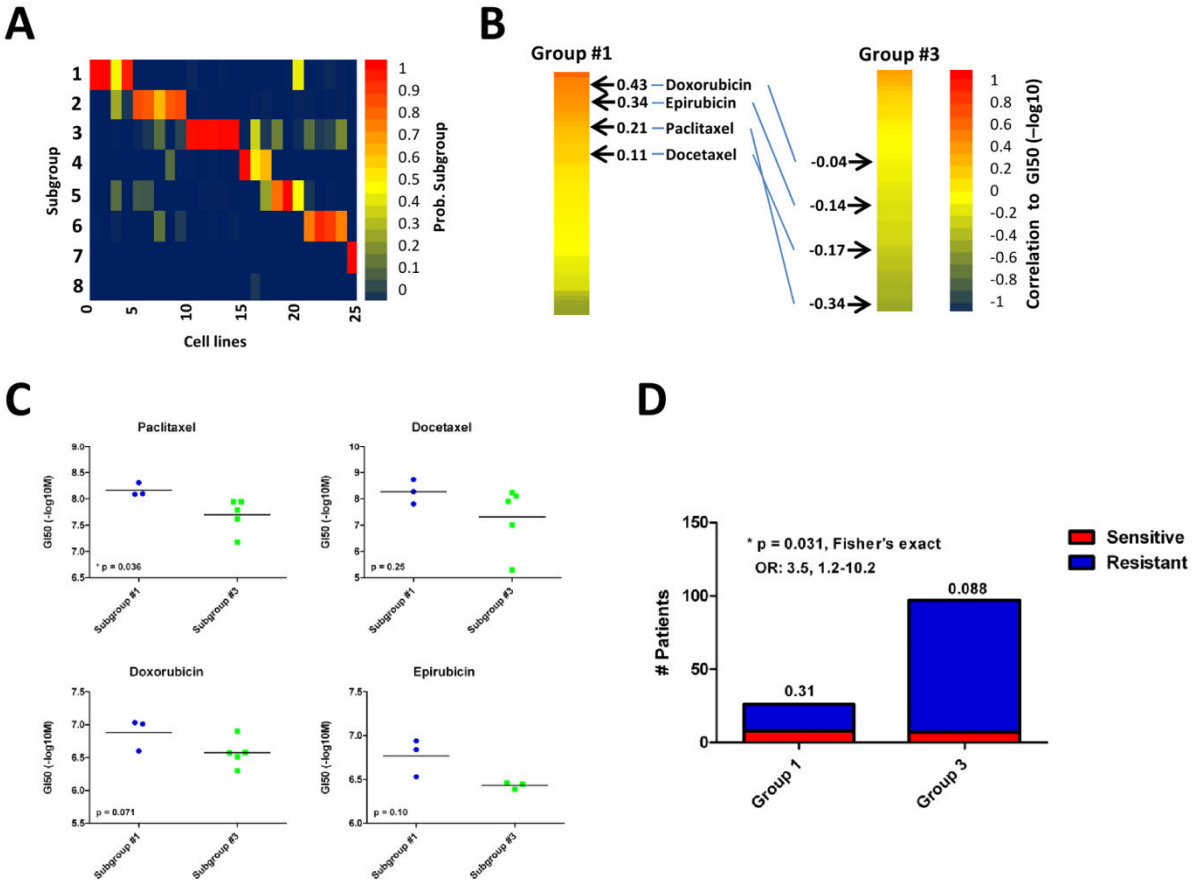
NCI-Nature curated pathways	***P***
Subgroup #1	
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	1.40E-04
Signaling events mediated by TCPTP	2.82E-04
ErbB1 downstream signaling	4.50E-04
IL6-mediated signaling events	4.86E-04
N-cadherin signaling events	6.73E-04
IL2 signaling events mediated by PI3K	8.97E-04
FGF signaling pathway	1.50E-03
Subroup #2	
Neurotrophic factor-mediated Trk receptor signaling	1.29E-03
p75(NTR)-mediated signaling	2.17E-03
PLK1 signaling events	3.03E-03
Degradation of beta catenin	2.28E-02
Hedgehog signaling events mediated by Gli proteins	2.34E-02
TGF-beta receptor signaling	3.23E-02
Subroup #3	
ErbB4 signaling events	5.63E-07
FGF signaling pathway	9.30E-06
IL6-mediated signaling events	3.36E-05
Regulation of nuclear SMAD2/3 signaling	9.78E-05
IL4-mediated signaling events	2.05E-04
Signaling events mediated by TCPTP	2.44E-04

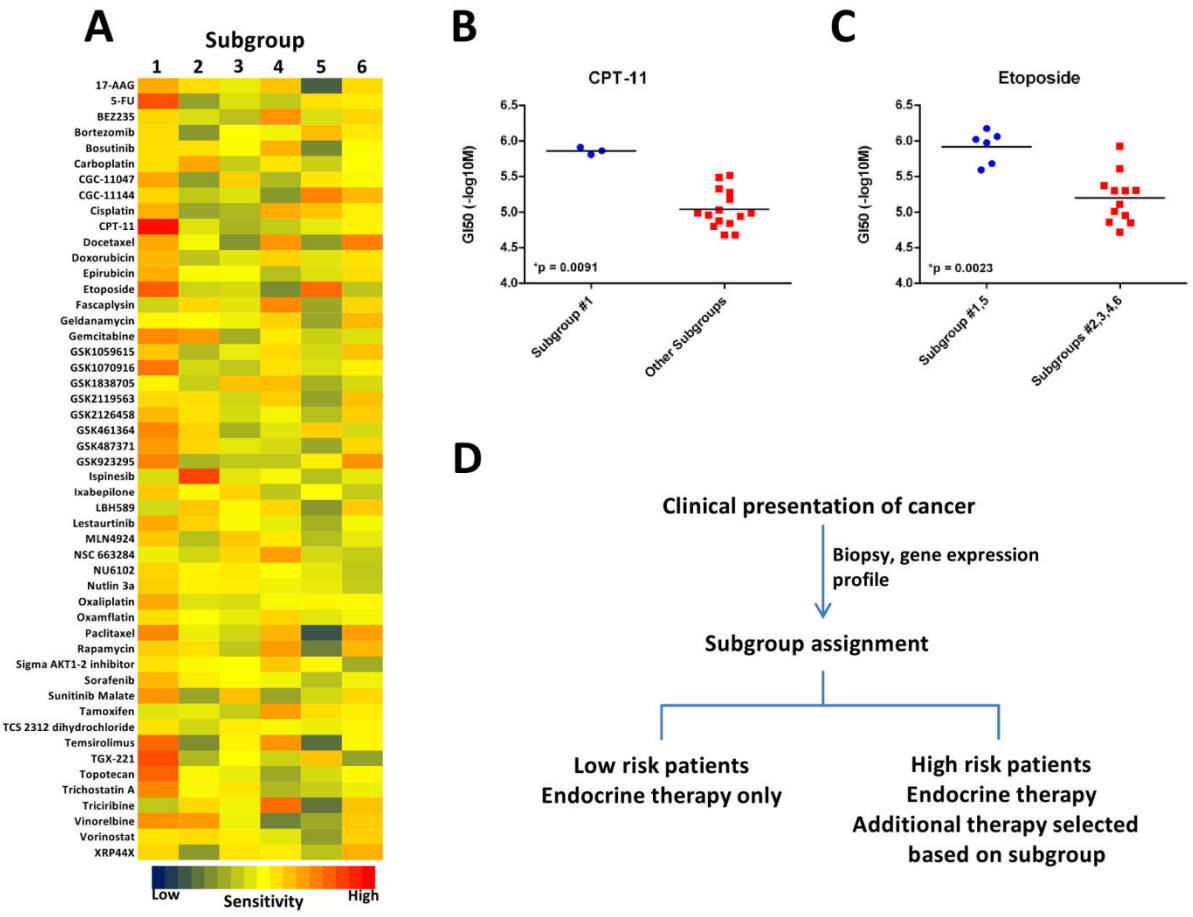
Subroup #4	
Signaling events mediated by HDAC Class I	2.44E-06
Signaling events mediated by HDAC Class II	7.79E-05
p73 transcription factor network	1.82E-04
ErbB1 downstream signaling	2.51E-04
Posttranslational regulation of adherens junction stability and dissassembly	4.07E-04
PDGFR-beta signaling pathway	1.15E-03
Subroup #5	
Regulation of Telomerase	4.33E-10
Regulation of RhoA activity	1.70E-06
p53 pathway	2.00E-05
Hypoxic and oxygen homeostasis regulation of HIF-1-alpha	2.25E-05
Regulation of CDC42 activity	4.12E-04
Coregulation of Androgen receptor activity	8.49E-04
Subroup #6	
Regulation of RhoA activity	4.63E-04
Arf6 signaling events	1.61E-03
Arf1 pathway	3.13E-03
TRAIL signaling pathway	6.98E-03
Alpha6 beta4 integrin-ligand interactions	1.03E-02
mTOR signaling pathway	1.55E-02
Subgroup #7	
CXCR4-mediated signaling events	7.92E-12
IL12-mediated signaling events	1.63E-10
Downstream signaling in naïve CD8+ T cells	4.25E-09
TCR signaling in naïve CD4+ T cells	2.67E-08

TCR signaling in naïve CD8+ T cells	3.28E-08
IFN-gamma pathway	3.48E-07
Subgroup #8	
Signaling events mediated by HDAC Class III	1.35E-13
mTOR signaling pathway	9.32E-07
Insulin-mediated glucose transport	1.85E-05
Regulation of nuclear beta catenin signaling and target gene transcription	2.58E-05
p38 signaling mediated by MAPKAP kinases	2.67E-05
Syndecan-4-mediated signaling events	3.65E-05


A
Pre/post treatment samples (GSE5462)
Maximize between/within variation
Estrogen independent genes
B
Between / Within score
Probe
Estrogen independent probes
Other
C
k= 2
k= 3
k= 4
k= 5
k= 6
k= 7
k= 8
k= 9
k= 10
D
PAM selected
Relative expression
E
% DMFS
Time (years)
* p = 0.03
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8


A
Subgroup
Breast Tumors
Prob. Subgroup
B
% DMFS
Time (years)
* p = 0.01
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
C
% DMFS
Time (years)
* p = 0.02
D
% DMFS
Time (years)
* p = 0.04
E
% DMFS
Time (years)
Group 3/6/7
Group 1/2/4/5/8
HR: 2.55, *p = 0.0037


A
Subgroup
Cell lines
Prob. Subgroup
B
Group #1
0.43 — Doxorubicin
0.34 — Epirubicin
0.21 — Paclitaxel
0.11 — Docetaxel
Group #3
-0.04
-0.14
-0.17
-0.34
Correlation to GI50 (−log10)
C
Paclitaxel
Docetaxel
Doxorubicin
Epirubicin
GI50 (-log10M)
* p = 0.036
p = 0.25
p = 0.071
p = 0.10
Subgroup #1
Subgroup #3
D
* p = 0.031, Fisher's exact
OR: 3.5, 1.2-10.2
Sensitive
Resistant
Patients
0.31
0.088
Group 1
Group 3


A
Subgroup
1 2 3 4 5 6
17-AAG
5-FU
BEZ235
Bortezomib
Bosutinib
Carboplatin
CGC-11047
CGC-11144
Cisplatin
CPT-11
Docetaxel
Doxorubicin
Epirubicin
Etoposide
Fascaplysin
Geldanamycin
Gemcitabine
GSK1059615
GSK1070916
GSK1838705
GSK2119563
GSK2126458
GSK461364
GSK487371
GSK923295
Ispinesib
Ixabepilone
LBH589
Lestaurtinib
MLN4924
NSC 663284
NU6102
Nutlin 3a
Oxaliplatin
Oxamflatin
Paclitaxel
Rapamycin
Sigma AKT1-2 inhibitor
Sorafenib
Sunitinib Malate
Tamoxifen
TCS 2312 dihydrochloride
Temsirolimus
TGX-221
Topotecan
Trichostatin A
Triciribine
Vinorelbine
Vorinostat
XRP44X
Low
Sensitivity
High
B
CPT-11
GI50 (-log10M)
6.5
6.0
5.5
5.0
4.5
4.0
*p = 0.0091
Subgroup #1
Other Subgroups
C
Etoposide
GI50 (-log10M)
6.5
6.0
5.5
5.0
4.5
4.0
*p = 0.0023
Subgroup #1,5
Subgroups #2,3,4,6
D
Clinical presentation of cancer
Biopsy, gene expression profile
Subgroup assignment
Low risk patients
Endocrine therapy only
High risk patients
Endocrine therapy
Additional therapy selected based on subgroup

CHAPTER 4

Introduction

Oncologists are faced with the challenging task of selecting the most effective therapies for individual cancer patients to achieve the best possible outcome. Indeed, the latter is the central goal of personalized cancer medicine. Many current studies focus on identifying similar gene signatures to guide the selection of appropriate chemotherapy regimens (Hatzis *et al.*; Hess *et al.*, 2006; Lee *et al.*; Rouzier *et al.*, 2005). However, gene signatures that predict tumor sensitivity or resistance to chemotherapy are not currently clinically available. The development and clinical implementation of gene signatures that predict response to commonly used chemotherapeutic agents could facilitate selecting the most efficacious therapeutic regimen given the molecular characteristics of an individual's tumor. Furthermore, therapy-predictive gene signatures could ensure patients do not receive ineffective and potentially deleterious chemotherapeutic regimens.

For this study we used complete pathological response (pCR) or residual disease (RD) as the clinical measurements to determine the sensitivity of a tumor to chemotherapy. In this fashion, patients in which no invasive or metastatic breast cancer can be detected following treatment are classified as having achieved complete pathological response (pCR), whereas patients that fail to achieve pCR are classified as having residual disease (RD). Importantly, neoadjuvant chemotherapy has been found to be as efficacious as chemotherapy given in the adjuvant setting, and patients who achieve complete pathological response after neoadjuvant intervention generally have an

excellent probability of experiencing long-term survival [13-15]. Taken together, these data suggest that response to neoadjuvant chemotherapy (pCR/RD) provides a relevant clinical model to develop and validate gene signature based predictors of breast tumor response to chemotherapy.

We sought to test whether *TOP2A* and *β-tubulin* transcript expression indices could predict response to commonly used chemotherapeutic agents, as the protein products of these genes represent the respective molecular targets of commonly used anthracycline- and taxane-related drugs (Abal *et al.*, 2003; Desmedt *et al.*; Mozzetti *et al.*, 2005). We hypothesized that such target-based expression indices would provide a biologically comprehensive measurement of either *TOP2A or β-tubulin* activity in a patient's tumor, and thus its likely dependence on either of these targets. These analyses establish an effective method for identifying predictive drug response signatures, and highlight the use of predictive gene signatures to guide the selection of anthracycline and taxane based chemotherapy regimens for breast cancer patients. Importantly, the expression indices reported here are independently predictive of response to chemotherapy in a multivariate model that also included age, nodal status, tumor grade and estrogen receptor status in a group of 488 patients treated with anthracycline and taxane based chemotherapy.

Manuscript #3

Title: A target based approach identifies genomic predictors of breast cancer patient response to chemotherapy

Authors: Robin M. Hallett[1], Gregory Pond[2] and John A. Hassell[1*]

1. Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5
2. Department of Oncology, Juravinski Hospital and Cancer Centre, 711 Concession Street, Hamilton, Ontario, L8V 1C3

*Correspondence and requests for materials should be addressed to J.A.H. (hassell@mcmaster.ca)

Acknowledgements

This work was generously supported by grants from the Canadian Stem Cell Network, the Ontario Institute for Cancer Research and the Canadian Breast Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We wish to acknowledgements helpful discussion from Dr. Anita Bane throughout the course of this work.

Author contributions

Robin Hallett – Conceived, planned, analyzed and performed most of the experiments in the paper, wrote the manuscript

Greg Pond – Performed statistical analyses for table 3

John Hassell – Helped write manuscript, provided critical feedback for project

Abstract

Background: The efficacy of chemotherapy regimens in breast cancer patients is variable and unpredictable. Whether individual patients either achieve long-term remission or suffer recurrence after therapy may be dictated by intrinsic properties of their breast tumors including genetic lesions and consequent aberrant transcriptional programs. Global gene expression profiling provides a powerful tool to identify such tumor-intrinsic transcriptional programs, whose analyses provide insight into the underlying biology of individual patient tumors. For example, multi-gene expression signatures have been identified that can predict the likelihood of disease reccurrence, and thus guide patient prognosis. Whereas such prognostic signatures are being introduced in the clinical setting, similar signatures that predict sensitivity or resistance to chemotherapy are not currently clinically available.

Results: Here, we present target based expression indices that predict breast tumor response to anthracycline and taxane based chemotherapy. Indeed, these signatures were independently predictive of chemotherapy response after adjusting for standard clinic-pathological variables such as age, grade, and estrogen receptor status in a cohort of 488 breast cancer patients treated with adriamycin and taxotere.

Conclusions: Importantly, our findings suggest the practicality of developing target based indices that predict response to therapeutics, as well as highlight the possibility of using gene signatures to guide the use of chemotherapy during treatment of breast cancer patients.

Results

TOP2A and β-tubulin expression are associated with complete pathological response in breast cancer patients treated with chemotherapy.

As proof-of-principle that target expression could be linked to chemotherapy response, we tested whether transcripts of *TOP2A* were associated with pCR in a relatively large (GSE21094, n=278, n=56 pCR, n=222 RD) cohort of breast cancer patients treated with a neoadjuvant chemotherapy regimen that contained an anthracycline (TFAC; paclitaxel [T], 5-FU [F], adriamycin [A] and cyclophosphamide [C]). *TOP2A* transcript levels were associated with increased response to chemotherapy (Fig. 1A, AUC: 0.61, p = 0.008), and the mean expression of *TOP2A* transcripts was higher in patients who showed a complete pathological response than those who displayed residual disease after treatment (Fig. 1B, RD: 122.5, pCR: 167.7, p = 0.04, t-test, Welch's correction). Importantly, these data suggested that elevated expression levels of *TOP2A* transcripts are associated with response to anthracycline based chemotherapy regimens, and are consistent with previous studies that link *TOP2A* expression with response to anthracycline therapy [19-21]. To extend these findings, we also tested whether *β-tubulin* expression was associated with pCR in a cohort of breast cancer patients treated with docetaxel. *β-tubulin* transcript levels were linked with increased response to docetaxel (Fig. 1C, AUC: 0.9, p= 0.001), and the mean expression of *β-tubulin* transcripts was higher in complete responders than those whose tumors did not achieve a complete response (Fig. 1D, RD: 2344, pCR: 1364, p < 0.0001, t-test).

Indeed, these findings are consistent with previous studies that link *β-tubulin* expression with sensitivity to taxane chemotherapy [22]. In total, these experiments suggest that target expression is associated with response to their chemotherapeutic agent counterparts.

The *TOP2A index is associated with complete pathological response in breast cancer patients treated with anthracyclines*

Our previous observations suggested the expression of transcripts encoding the protein targets of therapeutic agents was associated with response to these same therapies. To extend these findings, we calculated a *TOP2A* transcript expression index by identifying genes whose expression displayed either positive or negative correlation to the expression of *TOP2A* transcripts in 3 independent publicly available data sets (GSE2034, GSE7390, GSE6532) (Fig. 2A). In short, we measured the correlation of all probe sets to *TOP2A* expression and selected those that ranked among the top and bottom 1% of all probe sets in each data set. The latter approach identified 124 *TOP2A* associated probe sets, 86 that displayed positive association to *TOP2A* transcripts and 34 that were negatively associated with *TOP2A* transcript*s* (Fig. 2B, Supplementary Table 1). Using the same 278 patient cohort we employed previously, the *TOP2A* index was significantly associated with pCR (Fig. 2D, AUC: 0.73, $p < 0.0001$) and the *TOP2A* index scores were higher in patients who demonstrated a pCR than those who retained RD after treatment (Fig. 2E, RD: -0.15, pCR: 0.60, $p < 0.0001$, t-test,). Notably, the *TOP2A* index included genes linked to DNA repair, including *EXO1*[23], *ERCC6L*[24,

25], and *RAD51*[26], suggesting that the *TOP2A* index has a functional connection to the mechanisms of action of anthracycline drugs. To look for similarities between the *TOP2A* index and other predictors of chemotherapy response, we examined whether *TOP2A* index probes were also present within the DLDA30 predictor. Importantly, the DLDA30 is a validated gene expression based predictor of response to taxanes and anthracyclines, as well as to 5-fluorouracil and cyclophosphamide [9, 10, 12]. Interestingly, only one probe set from the *TOP2A* index also comprised an element in the DLDA30 predictor, suggesting minimal overlap between these two gene expression based predictors. Taken together, these data provide proof-of-principle that a target based expression index can predict response to small molecules that inhibit the activity of the given target under investigation. Importantly, we show here that the *TOP2A* index was associated with complete pathological response in breast cancer patients treated with an anthracycline-containing chemotherapy regimen in the neoadjuvant setting.

A β-tubulin index is *associated with complete pathological response in breast cancer patients treated with taxanes*

Given the development of the *TOP2A* index, we explored the possibility of identifying and developing a similar index to predict response to taxane drugs (Fig. 3A). In a similar fashion as reported above, we identified an expression index of *β-tubulin* associated transcripts. *β-tubulin* is the molecular target of taxane drugs, a commonly used class of chemotherapeutic drugs that includes docetaxel and paclitaxel (Taxotere® and Taxol® respectively)[27-29]. In short, we identified 42 *β-tubulin* transcript-

associated probe sets, 28 that displayed positive association with *β-tubulin* transcripts and 14 that were negatively associated with *β-tubulin* transcript levels (Fig. 3B). We evaluated the association between the *β-tubulin* index and response to docetaxel using a 14 patient (profiled in replicate, n=28, RD: 20, pCR: 8) cohort that was treated with neoadjuvant docetaxel (Fig. 3C). The *β-tubulin* index was associated with complete pathological response (Fig. 3D, AUC: 0.89, $p = 0.002$) and *β-tubulin* index scores were higher in patients who achieved complete pathological response than those who experienced residual disease after docetaxel therapy (Fig. 3E, RD: -0.5, pCR: 1.3, $p < 0.0001$, t-test). Interestingly, the *β-tubulin* index included genes linked to cytoskeleton processes, including *TPX2*[30], and *DBN1*[31], suggesting a linkage between the *β-tubulin* index and the mechanism of action of taxanes. As completed previously, we also looked for similarities between the *β-tubulin* index and the DLDA30 predictor. Notably, no probe sets from the *β-tubulin* index comprised an element within the DLDA30 predictor, suggesting minimal overlap between these two gene expression based predictors. In total, these data provide evidence that target based expression indices can predict response to neoadjuvant chemotherapy. In each case, the area under curve (AUC) was significantly greater than 0.5, confirming the predictive capacity of the indices, as well as the overall validity of the approach.

Combining indices is predictive of response to multi-agent chemotherapy

Neoadjuvant chemotherapy generally comprises multiple different chemotherapeutic agents. To determine whether we could combine our individual target

indices and predict response to multi-agent chemotherapy we tested the *TOP2A* and *β-tubulin* indices as a combination index with the 278- patient cohort (TFAC), as well as with two different cohorts of patients treated with neoadjuvant therapy comprising an anthracycline and a taxane (AT) (GSE25055 [n=310] & GSE25065 [n=198]). The individual *TOP2A* and *β-tubulin* indices were associated with response among the TFAC-treated patients, and patients who experienced pCR had significantly higher individual index scores than those who retained residual disease (Fig. 4A, $p < 0.0001$ both cases, t-test). After combining the indices for the 278 TFAC-treated patient cohort, we observed a significant association between the combined indices and tumor response (Fig. 4B, AUC: 0.76, $p < 0.0001$), and the combined index scores were significantly higher in responders (pCR) than non-responders (RD) (Fig. 4C, RD: -0.4, pCR: 1.7, $p < 0.0001$, t-test). Interestingly, the AUC value for the combined *TOP2A* and *β-tublin* index was nominally higher (AUC: 0.78) than that produced by the *TOP2A* index alone (Fig. 2D, AUC:0.73). We also tested the *TOP2A* and *β-tubulin* combination index on an additional two independent cohorts of patients treated with AT neoadjuvant chemotherapy (GSE25055 [n=310] & GSE25065 [n=198]). In the 310-patient cohort the individual *TOP2A* and *β-tubulin* index scores were significantly higher in responders than non-responders (Fig. 4D, $p < 0.0001$ both cases, t-test), the combination of the *TOP2A* and *β-tubulin* indices were associated with increased response to treatment (Fig. 4E, AUC: 0.76, $p < 0.0001$), and the *TOP2A/β-tubulin* combination index scores were higher in responders than non-responders (Fig. 4F, RD: -0.4, pCR: 1.6, $p < 0.0001$, t-test). We made similar observations in the 198-patient cohort, where individual *TOP2A* and *β-tubulin* index

scores were significantly higher in responders than in non-responders (Fig. 4G, $p < 0.0001$ both cases, t-test), and the combination index score was associated with complete pathological response (Fig. 4H, AUC: 0.75, $p < 0.0001$). Again, combined index scores were higher in responders than in non-responders (Fig. 4I, RD: -0.3, pCR: 1.3, $p < 0.0001$, t-test). Taken together these data reveal the feasibility of combining individual target indices to predict response to multi-agent chemotherapy regimens.

TOP2A and β-tubulin indices are more accurate than a similarly derived proliferation index

Previous work suggests that the prognostic/predictive power of many breast cancer gene expression signatures is derived from their capacity to measure proliferation [2, 4, 7, 32-36]. Hence, it is possible that the *TOP2A* and *β-tubulin* indices described here predict response based on their capacity to measure proliferation rather than providing a biologically relevant measurement of target. This is particularly relevant to the *TOP2A* index, as the protein product of the *TOP2A* gene is directly involved in DNA synthesis. To address this issue we first tested the capacity of the *TOP2A* index to predict response in the docetaxel-only treated cohort. In this fashion, the capacity of the *TOP2A* index to predict response to docetaxel would suggest that the predictive capacity of the *TOP2A* index is not target specific and may measure a more general phenomenom related to chemotherapy sensitivity, such as proliferation. Within this patient cohort, the *TOP2A* index was not significantly associated with patient response to docetaxel therapy (Fig. 5A&B, AUC: 0.68, p=0.14, RD: -0.26, pCR: 0.65, p=0.12, t-test). Indeed, these

results suggest that the predictive capacity of the *TOP2A* index is based on measurement of target rather than proliferation. We also generated a 'proliferation index' built using the same methodology as the *TOP2A* and *β-tubulin* indices, around the well characterized proliferation gene E2F1 [8, 37-41]. As expected, the E2F1 index was related to chemotherapy response in the 3 large cohorts of breast cancer patients 3 tested previously (GSE21094, GSE25055, GSE25065) (Supplementary Fig 1). Importantly, these results indicate the gene expression signatures that measure proliferation are associated with patient response to chemotherapy. However, the performance of the E2F1 predictor was inferior to either the *TOP2A* and *β-tubulin* predictors (Fig. 5C, *$p<0.05$, ANOVA, Tukeys). Taken together, these data suggest that the predictive capacity of the *TOP2A* index is specific to anthracycline drugs, and moreover, that the *TOP2A* and *β-tubulin* indices are superior to gene signatures that measure proliferation.

Comparison of the *TOP2A* and *β-tubulin* indices with clinico-pathologic parameters

Reported genomic tests for response to chemotherapy have generally failed to outperform the predictive capacity of standard clinic-pathological measurements [10]. In this regard we tested whether the *TOP2A/β-tubulin* combination index was related to tumor response after adjusting for standard clinical pathological variables. Combination of the two previously used data sets (GSE25055 & GSE25065, 20 patients discarded as pCR/RD data was unavailable) yielded a data set comprising 488 patients for the analysis. In a univariate analysis that included estrogen receptor status, tumor grade, nodal status, patient age, as well as the *TOP2A/β-tubulin* combination index score, only

estrogen receptor status (AUC: 0.68, $p < 0.001$), tumor grade (AUC: 0.69, $p < 0.001$) and the combination index score (Range [-5.829, 7.120], AUC: 0.76, $p < 0.001$) were found to be statistically significantly related to tumor response, whereas age and nodal status were not (Table 3). In a multivariate model (AUC: 0.78), the *TOP2A/β-tubulin* combination index score remained statistically significantly related to tumor response ($p < 0.001$) as well as estrogen receptor status ($p = 0.014$) and grade ($p = 0.016$) (Table 4). In the multivariate model the odds ratio for the *TOP2A/β-tubulin* index was 1.33, which indicated that for each unit increase (i.e. from -3 to -2, from 0 to 1, or from 4 to 5, etc) in the AT index, there was a 1.33 times increase in the odds of a patient experiencing complete pathological response. Importantly, these data suggest that the *TOP2A/β-tubulin* combination index score was related to patient response to chemotherapy even after adjusting for standard clinical-pathological variables.

Discussion

Here we describe the identification of *TOP2A* and *β-tubulin* transcript expression indices that predict complete pathological response to neoadjuvant chemotherapy regimens containing anthracycline and taxane drugs. Complete pathological response represents an appropriate clinical endpoint for these studies as patients who experience pCR also experience improved survival compared to those patients who retain RD [13-15]. Notably, *TOP2A or β-tubulin,* the respective targets of anthracycline and taxane drugs, have been linked to anthracycline and taxane response in previous studies, respectively [17, 18, 21, 42-44]. However, the expression of either of these genes has failed to become a useful clinical predictor of anthracycline or taxane response. We hypothesized that measurement of target-associated transcripts in a tumor sample might provide a more comprehensive measure of molecular target activity, and thus the tumor's likelihood of response to therapy. Indeed, based on the datasets explored for the studies presented here, this appears to be the case. Moreover, a combination index derived from the *TOP2A* and *β-tubulin* expression indices was statistically significantly related to pathological response in a multivariate model that also included age, nodal status, tumor grade and estrogen receptor status in a group of 488 patients treated with anthracycline and taxane based chemotherapy.

From a clinical standpoint, predicting response to anthracycline and taxane based chemotherapy may be useful to identify breast cancer patients who have a high likelihood of benefiting from such regimens. Conversely, patients predicted to be resistant to

anthracycline- and taxane-based chemotherapy may benefit from enrollment in clinical trials investigating the efficacy of novel treatments [45]. Many issues remain to be addressed to confirm the clinical utility of the *TOP2A* and *β-tubulin* indices. In this study our conclusions are based on the analysis of retrospective data, which limits its clinical value. Moreover, we did not establish or optimize a threshold that would serve to separate patients predicted likely to respond to therapy from those likely to be resistant. Additionally, we did not test the capacity of the *TOP2A* index to predict response to neoadjuvant chemotherapy that consisted of only of an anthracycline, suggesting that the *TOP2A* index may be predictive of general chemotherapy response. Athough we did observe that the *TOP2A* index was not predictive of patient response ot docetaxel. Based on our multivariate analysis, our data suggests the *TOP2A* and *β-tubulin* indices remain predictive even after adjusting for clinical parameters such as tumor grade and estrogen-receptor status, indicating that these indices likely have clinical value. Strictly speaking however, a true estimate of the usefulness of these indices would require a prospective clinical trial comparing randomly selected with index selected chemotherapy regimens.

An advantage of the approach taken here is our use of publicly available data, as well as the efficient use of patient samples for validation purposes. For example, the traditional approach for gene signature identification [2, 6, 7, 9, 46], commonly called the top-down approach, multiple datasets are required that comprise both tumor gene expression profiles as well as knowledge of the clinical variables under investigation, for the purposes of signature identification and subsequent independent validation. Other approaches, such as large-scale functional based RNA interference screens, have also

yielded predictive signatures, although these experiments are relatively labour intensive and expensive [47]. Here, we calculated target indices using datasets for which response to chemotherapy is not known. In this fashion, we maintained the independence of datasets for which response to neoadjuvant chemotherapy was measured as a clinical variable, thus maintaining the availability of multiple independent datasets for validation.

The identification of gene signatures that predict response to chemotherapy also have potential to offer new insights into the biology of breast tumors, particularly the transcriptional programs that govern therapy response. In this regard, it may be possible to identify molecular signaling pathways that either augment chemotherapy resistance or enhance sensitivity. Indeed, the latter strategy provides a rational approach to identifying new drug regimens, where a signaling pathway inhibitor/activator is included with the original chemotherapy regimen. In this fashion, tumors predicted to be therapy resistant might be rendered sensitive to the original therapy and treatment efficacy could be increased. Another important implication of this study is that it highlights the identification of target based expression indices as a means to predict response to therapeutics. For example, it might be possible to generate a target-based expression index for additional molecular targets, such as the *HER2/Neu* receptor tyrosine kinase, which is the molecular target of the humanized monoclonal antibody trastuzumab [1] as well as the small molecule Her2/Neu kinase inhibitor, lapatinib [48, 49]. Using such an approach, therapeutic response to the latter agents might then be predicted using transcriptional target based signatures. Indeed, this approach could be tested for multiple new experimental molecularly targeted therapies.

Methods

Patients and Samples

All data was publicly available and downloaded from the gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo/). Multiple discovery cohorts (GSE2034, GSE7390, GSE6532) were independently evaluated to determine target indices for *TOP2A, β-tubulin,* and *E2F1*. Together these cohorts comprised 811 patient tumor gene expression profiles derived from the Affymetrix U133A microarray platform (Table 1). Multiple validation cohorts were independently studied to test whether target indices were related to pathological response to neoadjuvant chemotherapy (GSE21094 (T/FAC), GSE22513 (T), GSE25055 & GSE25065 (AT). These cohorts comprised 800 patient tumor gene expression profiles from the Affymetrix U133A and Affymterix U133 Plus 2.0 microarray platforms (Table 1). The clinical characteristics of the validation cohorts are summarized in Table 2.

The raw intensity files (.CEL) comprising each dataset were download and normalised using the Robust Multichip Algorithm (RMA)to generate probe set intensities [50].

Identification of Target Related Genes

Target index genes were identified by their co-expression with either *TOP2A* (TOP2A, 201292_at*), β-tubulin* (TUBB, 212320_at), or *E2F1* (204947_at) based on a Pearson distance function [51]. We filtered these results such that only probe sets

appearing in the most and least 1% of co-expressed probe sets within each identification cohort were included in the target index. The final *TOP2A* index comprised 86 probe sets with positive and 38 probe sets with negative correlation to *TOP2A* transcript levels (Supplementary Table 1). The *β-tubulin* index comprised 28 probe sets with positive and 14 probe sets with negative correlation to *β-tubulin* transcript levels (Supplementary Table 2).

To evaluate the target index, the expression values for each probe set were transformed such that the mean and standard deviation were set to 0 and 1 in each dataset, respectively. A target index was calculated for each patient as follows:

$$\frac{\sum_{i \in P} x_i}{n_P} - \frac{\sum_{i \in N} x_j}{n_N}$$

Where x is the transformed expression, n is the number of probe sets, P is the set of probes with reported positive correlation to the target probe set, and N is the set of probes with reported negative correlation to the target probe set [46, 52].

Statistical Analysis

pCR or RD following treatment with neoadjuvant chemotherapy was used as the clinical endpoints for this study. The predictive capacities of the target indices were evaluated using receiver-operator characteristic curve (ROC) analysis and both univariate and multivariate logistic regression. T-tests were used to compare indices between responders and non-responders. Welch's correction was used when the variance of the

index was unequal in these two patient groups. All tests were two-sided and a p-value of 0.05 or less was considered statistically significant. ANOVA and Tukeys multitple comparison test was used to test for differences between multiple groups ($n > 2$), and p-values of 0.05 or less were considered significant.

References

Abal M, Andreu JM, Barasoain I (2003). Taxanes: microtubule and centrosome targets, and cell cycle dependent mechanisms of action. *Curr Cancer Drug Targets* **3:** 193-203.

Bear HD, Anderson S, Smith RE, Geyer CE, Jr., Mamounas EP, Fisher B *et al* (2006). Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer:National Surgical Adjuvant Breast and Bowel Project Protocol B-27. *J Clin Oncol* **24:** 2019-27.

Burgess DJ, Doles J, Zender L, Xue W, Ma B, McCombie WR *et al* (2008). Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc Natl Acad Sci U S A* **105:** 9053-8.

Dang C, Lin N, Moy B, Come S, Sugarman S, Morris P *et al* Dose-dense doxorubicin and cyclophosphamide followed by weekly paclitaxel with trastuzumab and lapatinib in HER2/neu-overexpressed/amplified breast cancer is not feasible because of excessive diarrhea. *J Clin Oncol* **28:** 2982-8.

Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J *et al* Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol* **29:** 1578-86.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP *et al* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286:** 531-7.

Hallett RM, Dvorkin-Gheva A, Anita B, Hassell JA (2012). A gene signature for predicting outcome in patients with basal-like breast cancer. *Scientific Reports* **2**.

Hallett RM, Dvorkin A, Gabardo CM, Hassell JA (2010). An algorithm to discover gene signatures with predictive potential. *J Exp Clin Cancer Res* **29:** 120.

Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A *et al* A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* **305:** 1873-81.

Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA *et al* (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24:** 4236-44.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al* (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249-64.

Kavallaris M, Kuo DY, Burkhart CA, Regl DL, Norris MD, Haber M *et al* (1997). Taxol-resistant epithelial ovarian tumors are associated with altered expression of specific beta-tubulin isotypes. *J Clin Invest* **100:** 1282-93.

Lee JK, Coutant C, Kim YC, Qi Y, Theodorescu D, Symmans WF *et al* Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer. *Clin Cancer Res* **16:** 711-8.

Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA *et al* (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* **26:** 1275-81.

Manfredi JJ, Horwitz SB (1984). Taxol: an antimitotic agent with a new mechanism of action. *Pharmacol Ther* **25:** 83-125.

Manfredi JJ, Parness J, Horwitz SB (1982). Taxol binds to cellular microtubules. *J Cell Biol* **94:** 688-96.

Moy B, Kirkpatrick P, Kar S, Goss P (2007). Lapatinib. *Nat Rev Drug Discov* **6:** 431-2.

Mozzetti S, Ferlini C, Concolino P, Filippetti F, Raspaglio G, Prislei S *et al* (2005). Class III beta-tubulin overexpression is a prominent mechanism of paclitaxel resistance in ovarian cancer patients. *Clin Cancer Res* **11:** 298-305.

Nikolenyi A, Sukosd F, Kaizer L, Csorgo E, Voros A, Uhercsak G *et al* Tumor topoisomerase II alpha status and response to anthracycline-based neoadjuvant chemotherapy in breast cancer. *Oncology* **80:** 269-77.

Parness J, Horwitz SB (1981). Taxol binds to polymerized tubulin in vitro. *J Cell Biol* **91:** 479-87.

Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K *et al* (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* **11:** 5678-85.

Scoggins JF, Ramsey SD A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. *J Natl Cancer Inst* **102:** 1371.

Slamon DJ, Press MF (2009). Alterations in the TOP2A and HER2 genes: association with adjuvant anthracycline sensitivity in human breast cancers. *J Natl Cancer Inst* **101:** 615-8.

Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J *et al* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* **98:** 262-72.

Tanner M, Isola J, Wiklund T, Erikstein B, Kellokumpu-Lehtinen P, Malmstrom P *et al* (2006). Topoisomerase IIalpha gene amplification predicts favorable treatment response to tailored and dose-escalated anthracycline-based adjuvant chemotherapy in HER-2/neu-amplified breast cancer: Scandinavian Breast Group Trial 9401. *J Clin Oncol* **24:** 2428-36.

Tommasi S, Mangia A, Lacalamita R, Bellizzi A, Fedele V, Chiriatti A *et al* (2007). Cytoskeleton and paclitaxel sensitivity in breast cancer: the role of beta-tubulins. *Int J Cancer* **120:** 2078-85.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L *et al* (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* **20:** 719-26.

Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B (2001). Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. *J Natl Cancer Inst Monogr***:** 96-102.

Yau C, Esserman L, Moore DH, Waldman F, Sninsky J, Benz CC A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *Breast Cancer Res* **12:** R85.

Table and Figure legends

Table 1. Summary of samples used to identify and validate target indices

Table 2. Characteristics of the validation cohorts

Table 3. Logistic regression analysis of the GSE25055 & GSE25065 validation set

Supplementary table 1. Probesets comprising *TOP2A* expression index

Supplementary table 2. Probesets comprising *β-tubulin* expression index

Figure 1. *TOP2A* and *β-tubulin* transcript levels predict response to anthracycline or taxane containing chemotherapy, respectively. A) ROC analysis of *TOP2A* transcript levels and response to TFAC neoadjuvant chemotherapy (AUC: 0.61, p = 0.008). **B)** Expression levels of *TOP2A* transcripts in patients who experienced pCR or had RD after neoadjuvant chemotherapy (p=0.04, t-test). **C)** ROC analysis of *β-tubulin* transcript levels and response to neoadjuvant docetaxel chemotherapy (AUC: 0.9, p = 0.001). **D)** Expression levels of *β-tubulin* transcripts in patients who experienced pCR or had RD after neoadjuvant docetaxel chemotherapy ($p < 0.0001$, t-test).

Figure 2. *TOP2A* index predicts response to anthracycline containing chemotherapy. A) Experimental strategy to identify the *TOP2A* index. **B)** *TOP2A* co-expressed probesets within the three discovery cohorts. **C)** Experimental strategy to validate the *TOP2A* index. **D)** ROC Analysis of *TOP2A* index scores and response to T/FAC neoadjuvant chemotherapy (AUC: 0.73, $p < 0.0001$). **E)** *TOP2A* index scores of patients who experienced pCR or had RD after neoadjuvant chemotherapy ($p < 0.0001$, t-test).

Figure 3. *β-tubulin* index predicts response to docetaxel chemotherapy. A) Experimental strategy to identify the *β-tubulin* index. **B)** *β-tubulin* co-expressed probesets within the three discovery cohorts. **C)** Experimental strategy to validate the *β-tubulin* index. **D)** ROC analysis of *β-tubulin* index scores and response to docetaxel neoadjuvant chemotherapy (AUC: 0.89, $p = 0.002$). **E)** *β-tubulin* index scores in patients that experienced pCR or had RD after neoadjuvant chemotherapy ($p < 0.0001$, t-test).

Figure 4. *TOP2A* and *β-tubulin* combination index predicts response to multi-agent anthracycline/taxane chemotherapy. A) Individual target index scores (A: *TOP2A* index, T: β-*tubulin* index) in patients who experienced complete pathological response or had residual disease after TFAC neoadjuvant chemotherapy (GSE21094, $p < 0.0001$ both cases, t-test). **B)** ROC analysis of combination index scores and response to TFAC neoadjuvant chemotherapy (AUC: 0.78, $p < 0.0001$). **C)** Combination index scores in patients who experienced pCR or had RD after neoadjuvant chemotherapy ($p < 0.0001$, t-test). **D)** Individual target index scores of patients who experienced pCR or had RD after AT neoadjuvant chemotherapy (GSE25055, $p < 0.0001$ both cases, t-test). **E)** ROC Analysis of combination index scores and response to AT neoadjuvant chemotherapy for GSE25055 (AUC: 0.78, $p < 0.0001$). **F)** Combination index scores of patients who experienced pCR or had RD after AT neoadjuvant chemotherapy ($p < 0.0001$, t-test). **G)** Individual target index scores of patients who experienced pCR or had RD after AT neoadjuvant chemotherapy (GSE25065, $p < 0.0001$ both cases, t-test). **H)** ROC analysis of combination index scores and response to AT neoadjuvant chemotherapy for GSE25065 (AUC: 0.75, $p < 0.0001$). **I)** Combination index scores of patients who

experienced complete pathological response or had residual disease after AT neoadjuvant chemotherapy ($p < 0.0001$, t-test).

Figure 5. *TOP2A and β-tubulin* indices are more accurate than a similarly derived proliferation index. **A)** ROC analusis of *TOP2A* index and patient response to neoadjuvant docetaxel therapy (p=0.14). **B)** *TOP2A* index scores of patients who experienced pCR or had RD after neoadjuvant docetaxel chemotherapy (p =0.14, t-test). **C)** The *TOP2A* and *β-tubulin* indices outperform an *E2F1*-derived proliferation at predicting response to neoadjuvant chemotherapy in multiple patient datasets (*$p < 0.05$, ANOVA, Tukey's test).

Supplementary Figure 1. The *E2F1* index is predictive of chemotherapy response in multiple datasets A) GSE21094 (n=278, TFAC). **B)** GSE25055 (n=310, AT). **C)** GSE25065(n=198,AT).

Table 1. Summary of samples used to identify and validate target indices

Characteristic	Discovery cohorts			Validation cohorts			
	GSE2034	**GSE7390**	**GSE6532**	**GSE21094**	**GSE25055**	**GSE25065**	**GSE22513**
Samples	286	198	327	278	310	198	14
Regimen	N/A	N/A	N/A	TFAC	AT	AT	T
Array type	U133A	U133A	U133A	U133A	U133A	U133A	U133 Plus 2.0
	Total arrays: 811			**Total arrays: 800**			

Abbreviations: TFAC: paclitaxel/docetaxel, 5-fluorouracil, doxorubicin, cyclophosphamide; AT: doxorubicin, paclitaxel/docetaxel; T: docetaxel

Table 2. Characteristics of the validation cohorts

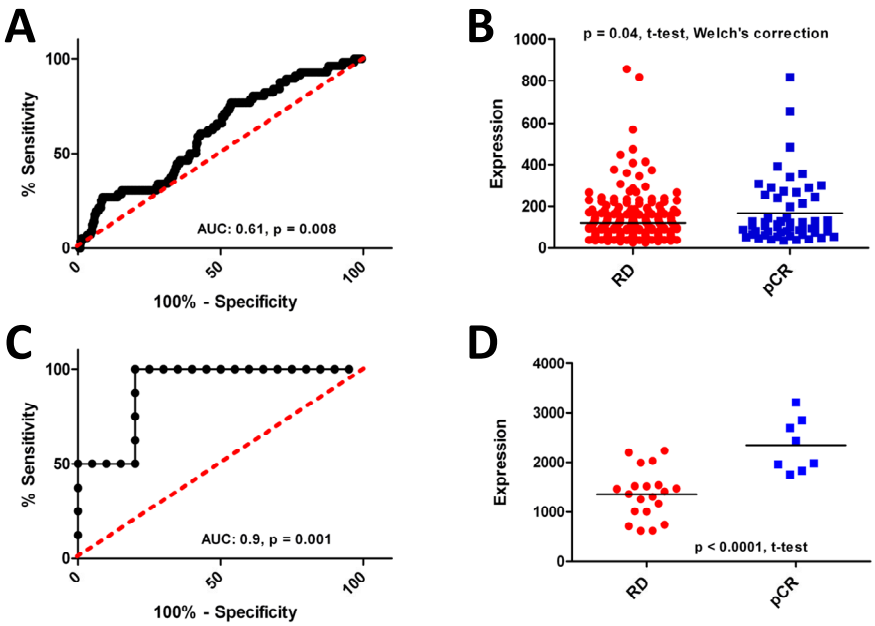
Characteristic	GSE21094		GSE25055		GSE25065		GSE22513	
	#	%	#	%	#	%	#	%
# patients	**278**		**310**		**198**		**14**	
Age, years							NA	
<=50	133	48	168	54	109	55		
>50	144	52	142	46	89	45		
NA	1	0						
Mean	52		50		49			
STD	11		10		11			
Nodal status	NA						NA	
Positive			223	72	128	65		
Negative			87	28	71	35		
NA								
Grade	NA						NA	
1			19	6	13	7		
2			117	38	63	32		
3			151	49	108	55		
NA			23	7	14	7		
Response								
pCR	56	20	58	19	42	21	4	29
RD	222	80	248	80	140	71	10	71
NA			4	1	16	8		
ER status							NA	
Positive	164	49	131	42	123	62		
Negative	114	41	174	56	74	37		
NA			5	2	1	1		

Abbreviations: STD: Standard deviation; pCR: complete pathological response; RD: Residual disease. ER: Estrogen receptor; NA: Not available

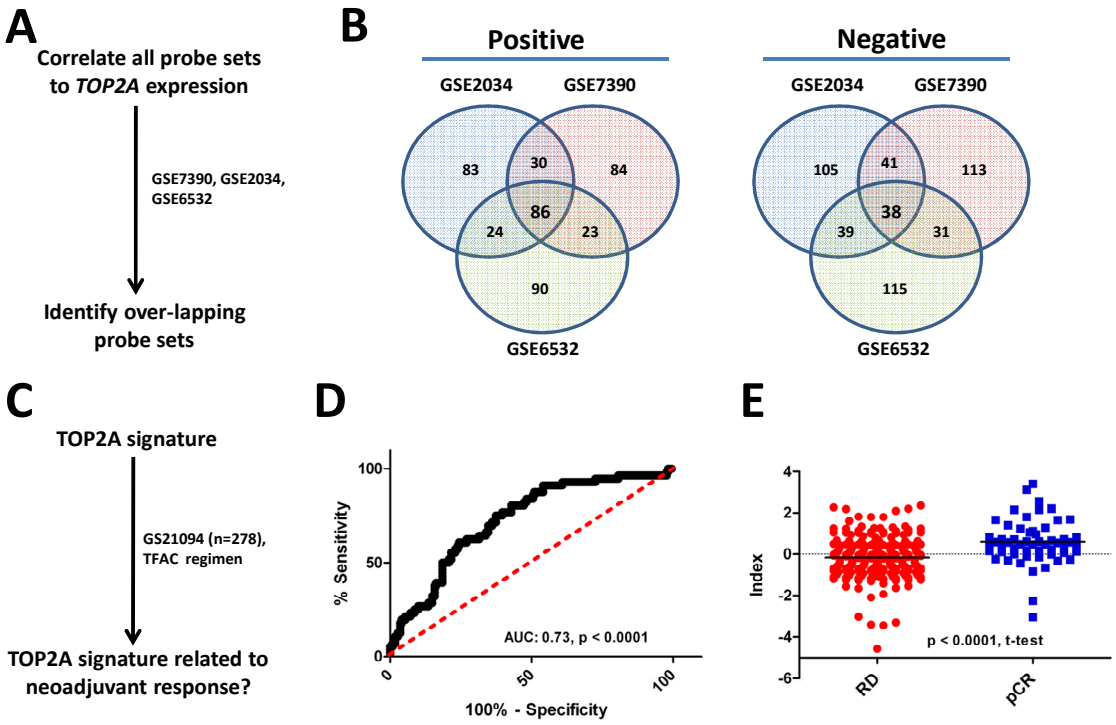
Table 3. Logistic regression analysis of the GSE25055 & GSE25065 validation set

Characteristic	Odds Ratio (95% CI)	AUC	p-value
	Univariate Analyses		
AT Index	1.57 (1.39, 1.77)	0.76	<0.001
Age (/10 years)	0.84 (0.68, 1.04)	0.55	0.10
ER Status Positive	0.23 (0.14, 0.37)	0.68	<0.001
Grade	5.28 (2.95, 9.46)	0.69	<0.001
Node Positive	1.16 (0.72, 1.88)	0.52	0.54
	Multivariate Analyses		
Age (/10 years)	0.89 (0.70, 1.13)	0.78	0.33
ER Status Positive	0.48 (0.27, 0.86)		0.014
Grade	2.24 (1.16, 4.33)		0.016
Node	0.95 (0.54, 1.67)		0.86
AT Index	1.33 (1.14, 1.55)		<0.001

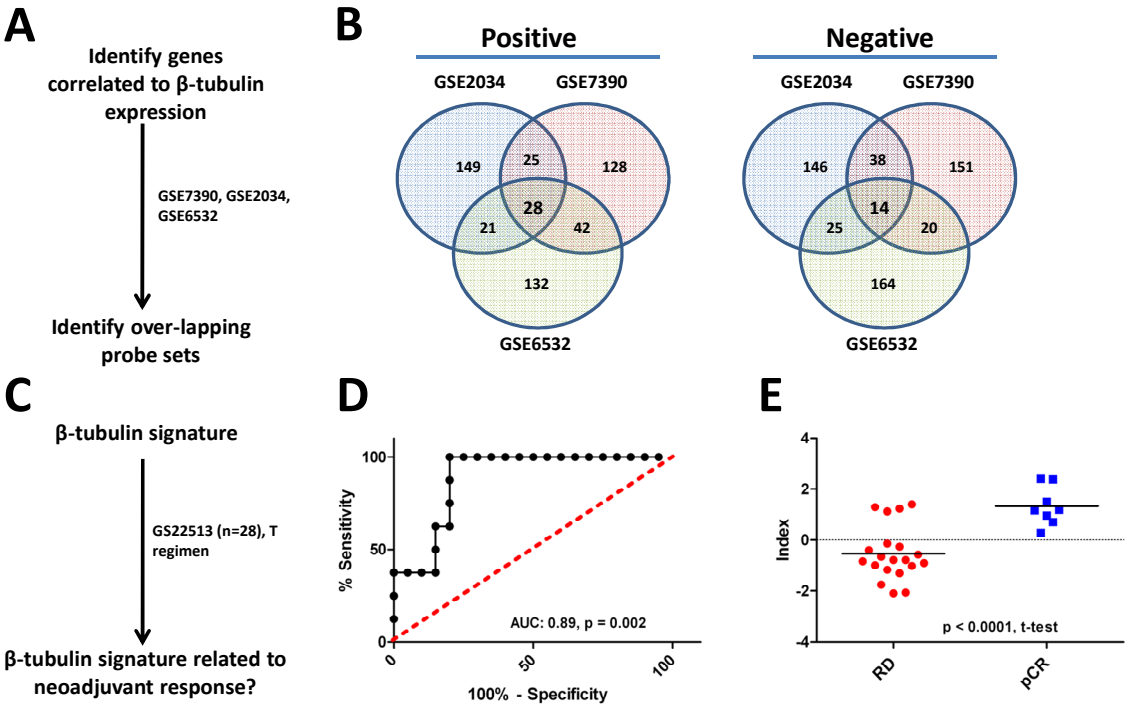
Hallett, *et al.* Figure 1



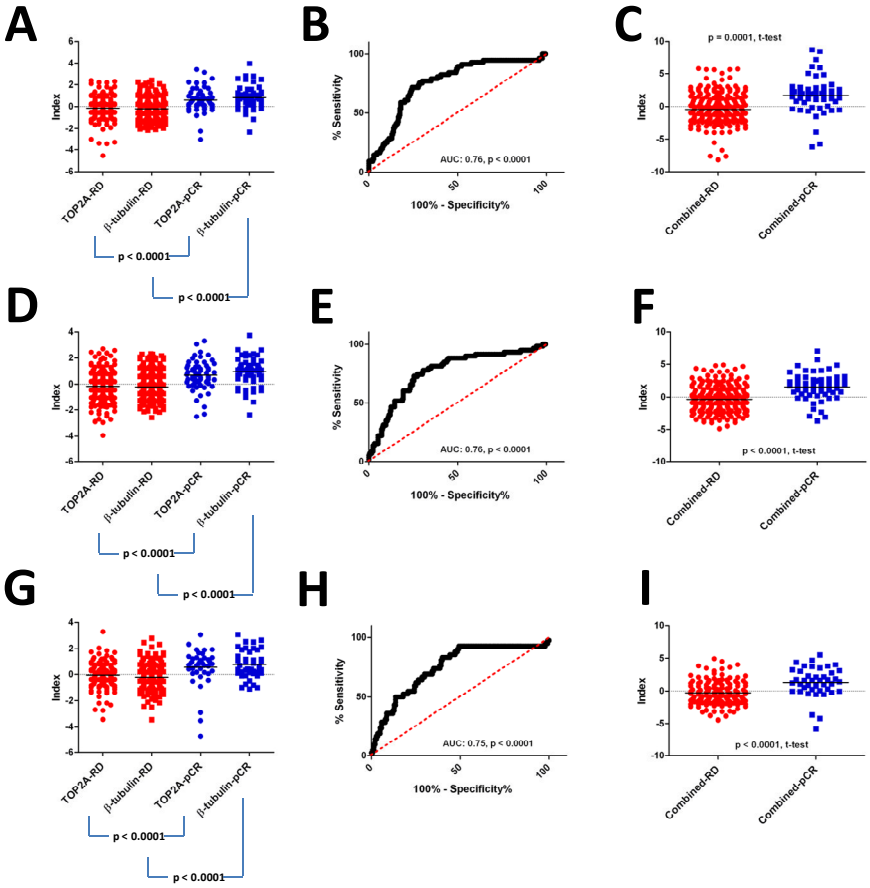
Hallett, *et al.* Figure 2



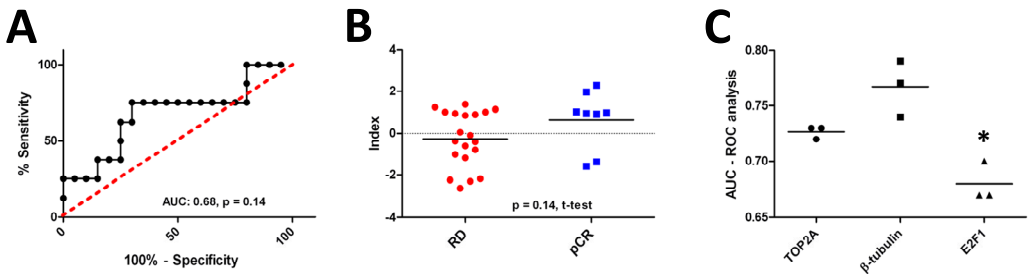
Hallett, *et al.* Figure 3



Hallett, *et al.* Figure 4



Hallett, *et al.* Figure 5



Supplementary table 1. Probesets comprising *TOP2A* expression index

TOP2A index		
Correlation	Name	Description
+	201292_at	topoisomerase (DNA) II alpha 170kDa, TOP2A
+	201291_s_at	topoisomerase (DNA) II alpha 170kDa, TOP2A
+	209408_at	kinesin family member 2C, KIF2C
+	211519_s_at	kinesin family member 2C, KIF2C
+	206364_at	kinesin family member 14, KIF14
+	204825_at	maternal embryonic leucine zipper kinase, MELK
+	218542_at	centrosomal protein 55kDa, CEP55
+	204444_at	kinesin family member 11, KIF11
+	202870_s_at	cell division cycle 20 homolog (S. cerevisiae), CDC20
+	219306_at	kinesin family member 15, KIF15
+	207165_at	hyaluronan-mediated motility receptor (RHAMM), HMMR
+	218355_at	kinesin family member 4A, KIF4A
+	204641_at	NIMA (never in mitosis gene a)-related kinase 2, NEK2
+	202954_at	ubiquitin-conjugating enzyme E2C, UBE2C
+	203213_at	cell division cycle 2, G1 to S and G2 to M, CDC2
+	218009_s_at	protein regulator of cytokinesis 1, PRC1
+	204318_s_at	G-2 and S-phase expressed 1, GTSE1
+	212949_at	non-SMC condensin I complex, subunit H, NCAPH
+	210559_s_at	cell division cycle 2, G1 to S and G2 to M, CDC2
+	218115_at	ASF1 anti-silencing function 1 homolog B (S. cerevisiae), ASF1B
+	218755_at	kinesin family member 20A, KIF20A
+	204267_x_at	protein kinase, membrane associated tyrosine/threonine 1, PKMYT1
+	209709_s_at	hyaluronan-mediated motility receptor (RHAMM), HMMR
+	212247_at	nucleoporin 205kDa, NUP205
+	222077_s_at	Rac GTPase activating protein 1, RACGAP1
+	210052_s_at	TPX2, microtubule-associated, homolog (Xenopus laevis), TPX2
+	219918_s_at	asp (abnormal spindle) homolog, microcephaly associated (Drosophila), ASPM
+	201088_at	karyopherin alpha 2 (RAG cohort 1, importin alpha 1), LOC728860
+	204026_s_at	ZW10 interactor, ZWINT
+	204092_s_at	aurora kinase A, AURKA
+	209714_s_at	cyclin-dependent kinase inhibitor 3, CDKN3
+	204603_at	exonuclease 1, EXO1
+	203755_at	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast), BUB1B
+	209642_at	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast), BUB1

+	203554_x_at	pituitary tumor-transforming 1, PTTG1
+	219588_s_at	non-SMC condensin II complex, subunit G2, NCAPG2
+	218726_at	Holliday junction recognition protein, HJURP
+	218039_at	nucleolar and spindle associated protein 1, NUSAP1
+	206316_s_at	kinetochore associated 1, KNTC1
+	203214_x_at	cell division cycle 2, G1 to S and G2 to M, CDC2
+	219650_at	excision repair cross-complementing rodent repair deficiency, ERCC6L
+	207828_s_at	centromere protein F, 350/400ka (mitosin), CENPF
+	208511_at	pituitary tumor-transforming 3, PTTG3
+	38158_at	extra spindle pole bodies homolog 1 (S. cerevisiae), ESPL1
+	202503_s_at	KIAA0101, KIAA0101
+	204126_s_at	CDC45 cell division cycle 45-like (S. cerevisiae), CDC45L
+	205046_at	centromere protein E, 312kDa, CENPE
+	202705_at	cyclin B2, CCNB2
+	219556_at	chromosome 16 open reading frame 59, C16orf59
+	222039_at	kinesin family member 18B, KIF18B
+	204170_s_at	CDC28 protein kinase regulatory subunit 2, CKS2
+	214710_s_at	cyclin B1, CCNB1
+	213226_at	cyclin A2, CCNA2
+	208808_s_at	high-mobility group box 2, HMGB2
+	221258_s_at	kinesin family member 18A, KIF18A
+	203046_s_at	timeless homolog (Drosophila), TIMELESS
+	208079_s_at	aurora kinase A, AURKA
+	209464_at	aurora kinase B, AURKB
+	204887_s_at	polo-like kinase 4 (Drosophila), PLK4
+	202589_at	thymidylate synthetase, TYMS
+	213008_at	Fanconi anemia, complementation group I, FANCI
+	204817_at	extra spindle pole bodies homolog 1 (S. cerevisiae), ESPL1
+	221520_s_at	cell division cycle associated 8, CDCA8
+	205167_s_at	cell division cycle 25 homolog C (S. pombe), CDC25C
+	203362_s_at	MAD2 mitotic arrest deficient-like 1 (yeast), MAD2L1
+	204822_at	TTK protein kinase, TTK
+	204649_at	trophinin associated protein (tastin), TROAP
+	204886_at	polo-like kinase 4 (Drosophila), PLK4
+	205339_at	SCL/TAL1 interrupting locus, STIL
+	218883_s_at	MLF1 interacting protein, MLF1IP
+	204962_s_at	centromere protein A, CENPA
+	208795_s_at	minichromosome maintenance complex component 7, MCM7
+	209891_at	SPC25, NDC80 kinetochore complex component, homolog (S. cerevisiae), SPC25
+	218308_at	transforming, acidic coiled-coil containing protein 3, TACC3

+	201896_s_at	proline/serine-rich coiled-coil 1, PSRC1
+	210821_x_at	centromere protein A, CENPA
+	203432_at	thymopoietin, TMPO
+	219000_s_at	defective in sister chromatid cohesion 1 homolog (S. cerevisiae), DSCC1
+	219990_at	E2F transcription factor 8, E2F8
+	212020_s_at	antigen identified by monoclonal antibody Ki-67, MKI67
+	205733_at	Bloom syndrome, BLM
+	205024_s_at	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae), RAD51
+	211762_s_at	karyopherin alpha 2 (RAG cohort 1, importin alpha 1), LOC728860
+	201930_at	minichromosome maintenance complex component 6, MCM6
+	221591_s_at	family with sequence similarity 64, member A, FAM64A
+	201202_at	proliferating cell nuclear antigen, PCNA
-	205116_at	laminin, alpha 2 (merosin, congenital muscular dystrophy), LAMA2
-	214927_at	integrin, beta-like 1 (with EGF-like repeat domains), ITGBL1
-	214761_at	zinc finger protein 423, ZNF423
-	216264_s_at	laminin, beta 2 (laminin S), LAMB2
-	218418_s_at	KN motif and ankyrin repeat domains 2, KANK2
-	202177_at	growth arrest-specific 6 /// similar to growth arrest-specific 6, GAS6 /// LOC100133684
-	219304_s_at	platelet derived growth factor D, PDGFD
-	202994_s_at	fibulin 1, FBLN1
-	218483_s_at	chromosome 11 open reading frame 60, C11orf60
-	209550_at	necdin homolog (mouse), NDN
-	208451_s_at	complement component 4A (Rodgers blood group), C4B
-	212494_at	tensin like C1 domain containing phosphatase (tensin 2), TENC1
-	213519_s_at	laminin, alpha 2 (merosin, congenital muscular dystrophy), LAMA2
-	212695_at	cryptochrome 2 (photolyase-like), CRY2
-	200811_at	cold inducible RNA binding protein, CIRBP
-	219993_at	SRY (sex determining region Y)-box 17, SOX17
-	212869_x_at	tumor protein, translationally-controlled 1, TPT1
-	214428_x_at	complement component 4A (Rodgers blood group), C4B
-	201591_s_at	nischarin, NISCH
-	206580_s_at	EGF-containing fibulin-like extracellular matrix protein 2, EFEMP2
-	206481_s_at	LIM domain binding 2, LDB2
-	214724_at	DIX domain containing 1, DIXDC1
-	201185_at	HtrA serine peptidase 1, HTRA1
-	209335_at	decorin, DCN
-	216520_s_at	tumor protein, translationally-controlled 1, TPT1
-	211896_s_at	decorin, DCN
-	213675_at	CDNA FLJ25106 fis, clone CBR01467, ---
-	219213_at	junctional adhesion molecule 2, JAM2

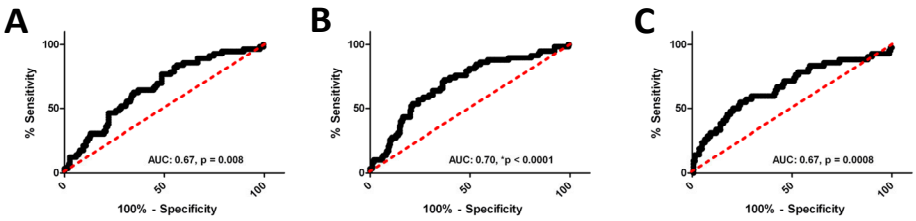
-	209687_at	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1), CXCL12
-	216840_s_at	laminin, alpha 2 (merosin, congenital muscular dystrophy), LAMA2
-	217807_s_at	glioma tumor suppressor candidate region gene 2, GLTSCR2
-	219315_s_at	transmembrane protein 204, TMEM204
-	202192_s_at	growth arrest-specific 7, GAS7
-	201893_x_at	decorin, DCN
-	212419_at	chromosome 10 open reading frame 56, C10orf56
-	202808_at	chromosome 10 open reading frame 26, C10orf26
-	212423_at	chromosome 10 open reading frame 56, C10orf56
-	200795_at	SPARC-like 1 (mast9, hevin), SPARCL1

Supplementary table 2. Probesets comprising *β-tubulin* expression index

B-tubulin index		
Correlation	Name	Description
+	212320_at	tubulin, beta, TUBB
+	209026_x_at	tubulin, beta, TUBB
+	200045_at	ATP-binding cassette, sub-family F (GCN20), member 1, ABCF1
+	210983_s_at	minichromosome maintenance complex component 7, MCM7
+	211714_x_at	tubulin, beta, TUBB
+	200934_at	DEK oncogene (DNA binding), DEK
+	208628_s_at	Y box binding protein 1, YBX1
+	209449_at	LSM2 homolog, U6 small nuclear RNA associated (S. cerevisiae), LSM2
+	208627_s_at	Y box binding protein 1, YBX1
+	200693_at	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, YWHAQ
+	201231_s_at	enolase 1, (alpha), ENO1
+	211519_s_at	kinesin family member 2C, KIF2C
+	200783_s_at	stathmin 1/oncoprotein 18, STMN1
+	210052_s_at	TPX2, microtubule-associated, homolog (Xenopus laevis), TPX2
+	203061_s_at	mediator of DNA damage checkpoint 1, MDC1
+	212021_s_at	antigen identified by monoclonal antibody Ki-67, MKI67
+	201797_s_at	valyl-tRNA synthetase, VARS
+	202589_at	thymidylate synthetase, TYMS
+	206074_s_at	high mobility group AT-hook 1, HMGA1
+	217025_s_at	drebrin 1, DBN1
+	209464_at	aurora kinase B, AURKB
+	201755_at	minichromosome maintenance complex component 5, MCM5
+	202326_at	euchromatic histone-lysine N-methyltransferase 2, EHMT2
+	201342_at	small nuclear ribonucleoprotein polypeptide C, SNRPC
+	221520_s_at	cell division cycle associated 8, CDCA8
+	204768_s_at	flap structure-specific endonuclease 1, FEN1
+	37425_g_at	coiled-coil alpha-helical rod protein 1, CCHCR1
+	214383_x_at	kelch domain containing 3, KLHDC3
-	201978_s_at	KIAA0141, KIAA0141
-	200711_s_at	S-phase kinase-associated protein 1, SKP1
-	218146_at	glycosyltransferase 8 domain containing 1, GLT8D1
-	218437_s_at	leucine zipper transcription factor-like 1, LZTFL1
-	201924_at	AF4/FMR2 family, member 1, AFF1
-	203538_at	calcium modulating ligand, CAMLG
-	205074_at	solute carrier family 22 (organic cation/carnitine transporter), member

		5, SLC22A5
-	200810_s_at	cold inducible RNA binding protein, CIRBP
-	214151_s_at	cell cycle progression 1, CCPG1
-	212510_at	glycerol-3-phosphate dehydrogenase 1-like, GPD1L
-	214164_x_at	carbonic anhydrase XII, CA12
-	218211_s_at	melanophilin, MLPH
-	213063_at	zinc finger CCCH-type containing 14, ZC3H14
-	209604_s_at	GATA binding protein 3, GATA3

Hallett, *et al.* Supplementary Figure 1



CHAPTER 5

Introduction

Modern chemotherapy began with the treatment of non-Hodgkin's lymphoma with nitrogen mustard, providing the first evidence that the systemic administration of drugs to cancer patients could induce tumor regression and prolong survival (Gilman, 1963). Later, the development of combination chemotherapy led to curative therapy for childhood acute lymphoblastic leukemia (ALL), as well as various types of lymphoma (Chabner and Roberts, 2005; Devita *et al.*, 1970). Numerous other advances have also been achieved in the development and application of chemotherapeutic agents, such as the discovery of cisplatin, which was later used to cure testicular tumors (Bosl *et al.*, 1986), and the development of anthracyclines that inhibit topoisomerase II (Minocha and Long, 1984). Despite these advances, chemotherapy generally fails to provide long-term cure, especially in the context of advanced disease (Breathnach *et al.*, 2001; Chabner and Roberts, 2005; Herbst *et al.*, 2006).

In attempt to improve response to treatment, recent work has focused on identifying molecular features of tumors that can be used to predict response to commonly used chemotherapeutic regimens (Ayers *et al.*, 2004; Hatzis *et al.*, 2011; Hess *et al.*, 2006; Potti and Nevins, 2008; Staunton *et al.*, 2001). In this fashion it is possible to select the most appropriate chemotherapeutic regimen to achieve maximal pathological response. For example, in the absence of patient selection, trastuzumab is clinically beneficial in slightly less than 10% of breast cancer patients. However, when patients are selected for trastuzumab therapy based on ERBB2/HER2 amplification, the response rate

rises dramatically to 35-50% (Vogel *et al.*, 2002). Another alternative for improving treatment response involves the use of sensitizing compounds to overcome chemotherapeutic resistance. Indeed, this is the rationale underpinning the development of poly-ADP ribose polymerase (PARP) inhibitors to treat cancer(Comen and Robson, 2010; Liang and Tan, 2010; Lord and Ashworth, 2008). More recent data suggests that treatment failure may be due to a minor population of tumor cells, termed cancer stem cells or tumor-initiating cells (TIC), which resist both radiation therapy and chemotherapy. Conventional therapies eradicate the proliferating descendants of TIC, which comprise the vast majority of cells populating tumors. Consequently tumors regress but often recur. Hence new therapies are needed that target TIC as well as the bulk tumor cells to provide durable breast cancer remission.

In this manuscript our goal was to identify novel sensitizing compounds to commonly used chemotherapy agents. To this end, we derived multiple transcriptional models of chemotherapy resistance. Importantly, these models also showed transcriptional connectivity to the gene expression changes induced by the small molecule adiphenine. Moreover, adiphenine sensitized multiple breast tumor cell lines to the cytotoxic effects of doxorubicin, a commonly used chemotherapeutic agent. Finally, additional transcriptional and experimental analyses suggested the chemotherapy-sensitizing effects of adiphenine may be linked to its capacity to target breast TIC.

Manuscript #4

Title: Use of genomic predictors to identify small molecule sensitizers to chemotherapy

Authors: Hallett, R.M.[1] Bering, G.[1] Liyanange, S.[1] Nixon, A.M.L.[1] Okolowsky, A.[1] Kent, J.[1] Gabardo, C.M.[1] Giacomelli, A.,O.[1] and Hassell, J. A[1*]

1. Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5

*Correspondence and requests for materials should be addressed to J.A.H. (hassell@mcmaster.ca)

.

Acknowledgements

This work was generously supported by grants from the Canadian Stem Cell Network, the Ontario Institute for Cancer Research and the Canadian Breast Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Robin Hallett – Conceived, planned, analyzed and performed most of the experiments in the paper, wrote the manuscript

Gursharn Bering, Sanduni Liyanage, Allison Nixon, Nadia Okolowsky, Jennifer Kent, Chrisitine Gabardo – Were project students that worked on various aspects of this project with Robin Hallett

Andrew Giacomelli – Provided technical expertise to the design of the alamar blue reduction assay

John Hassell – Helped write manuscript, provided critical feedback for project

Abstract

Despite advances in cancer detection and treatment, patients still generally fail to achieve long-term cure, especially in the context of advanced disease. Recent findings suggest that the latter may be the result of the persistence of a minor population of tumor cells, termed cancer stem cells or tumor-initiating cells (TIC), which resist current conventional anti-cancer therapies. In attempt to overcome chemotherapy resistance in breast cancer we used publicly available gene expression data to derive breast cancer patient- and breast tumor cell line-transcriptional models of chemotherapy resistance. We next used these models to mine the Connectivity Mapping to identify candidate compounds that overcome chemotherapy resistance. Both the patient and cell line-derived gene signatures identified adiphenine as a candidate compound that may abrogate chemo-resistance. Notably, adiphenine and analogs of adiphenine sensitized multiple breast tumor cell lines to the cytotoxic effects of doxorubicin, a commonly used chemotherapeutic agent. Additional transcriptional analysis of adiphenine regulated genes suggested that adiphenine targeted TIC, possibly accounting for its chemosensitizing properties. Indeed, adiphenine and select analogs blocked sphere formation, a surrogate test for TIC self-renewal, in several breast tumor cell lines. Taken together, our data suggest that adiphenine targets TICs and increases the susceptibility of breast tumor cell populations to the cytotoxic effects of chemotherapy. Moreover, these results highlight the opportunity for rational drug regimen design, where a sensitizing/anti-TIC agent is included with standard agents to improve clinical responses anti-cancer therapies.

Results

A gene signature predictor of breast cancer patient response to chemotherapy.

The observation that breast tumors show highly varied response to chemotherapy(Ayers *et al.*, 2004; Desmedt *et al.*, 2011; Hatzis *et al.*, 2011; Lee *et al.*) prompted us to identify patterns in gene expression associated with this phenomenon and subsequently build a transcriptional model of breast tumor chemotherapy response. To this end, we downloaded the gene expression profiles of 310 breast tumors for which response to neoadjuvant anthracycline and taxane therapy was measured (GSE2055) (Hatzis *et al.*, 2011). We separated this dataset into ER+ (n=172), and ER- (n=128) subgroups and used Bayesian probabilistic models (West *et al.*, 2001) to identify probe sets whose expression was associated with either complete pathological response (pCR) or residual disease (RD) after chemotherapy, for both ER+/ER- subgroups (Fig 1A). We then selected probe sets, which were associated with response in both ER+ and ER-negative patients. By first subdividing these patients into ER+/ER- subgroups, we avoided identifying probe sets that were simply associated with ER status, a potential confounding factor given the strong relationship between ER expression and response to chemotherapy. Importantly, this analysis returned 54 probe sets whose expression was associated with pCR, and 13 whose expression was associated with RD, thus providing a 67 probe set predictor of chemotherapy response (Fig. 1B&C). Within the training corhort (GSE25055, n=310), and two additional validation cohorts (GSE25065, n=198, GSE21094, n=278), the 67 probe sets were significantly predictive of chemotherapy response (Fig. 1D, GSE25055 AUC: 0.83, *p<0.001, GSE25065 AUC: 0.74, *p<0.0001,

GSE21094 AUC: 0.81, *p<0.0001). Notably, using a prospectively defined cut-off (signature score yielding 80% sensitivity in training cohort) to separate patients into responder/non-responder groups, the odds of responding were significantly greater in the predicted response groups than the predicted non-response groups. (Fig. 1D, GSE25055 Odds Ratio (OR): 12.3, GSE25065 OR: 4.8, GSE21094 OR: 9.5). Taken together, these data suggest that the 67 probe set predictor provides a robust transcriptional model of breast tumor response to chemotherapy.

Development of a gene signature predictor of breast tumor cell line response to chemotherapy.

To develop a gene signature predictor of breast tumor cell line response to chemotherapy, T) we used a methodology similar to that of previous studies(Staunton *et al.*, 2001). We assembled a panel of breast tumor cell lines and measured the response of each cell line to treatment with doxorubicin (Fig. 2A). In short, we seeded cells into 384 plates and incubated them for 48 hours in the presence of increasing concentrations of doxorubicin. To assess the cytotoxicity of doxorubicin, we completed alamarBlue reduction assays and measured alamarBlue reduction after 48 hours. Interestingly, we observed that breast tumor cell line response to doxorubicin was highly variable, where some cell lines were completely killed at a 100nM dose after 48hours, whereas others were relatively unscathed after the same treatment (Fig. 2B). Using the most (HCC1954, HCC1569, BT549, and SKBR3) and least (MCF7, MDA-MB-231, MDA-MB-453, and HS578T) resistant cell lines as well as their associated gene expression profiles, we identified probe sets whose expression was related to doxorubicin sensitivity/resistance

and used this data to build a model that could predict cell line sensitivity/resistance to doxorubicin. In short, the gene expression profiles of multiple breast tumor cell lines were downloaded (E-TABM-157) (Neve *et al.*, 2006) and we used Bayesian probalistic models to assess the association of each probe set with resistance/sensitivity to doxorubicin in the cell lines (Hallett, 2010) (Fig. 2C). In this fashion, we identified the top 250 probe sets for classification of cell lines on the basis of their response to doxorubicin (Fig. 2D). Importantly, leave-one-out cross-validation supports the capacity of these probe sets to accurately identify doxorubicin-resistant/sensitive cell lines (Fig. 2E, Resistant lines: 0.73, Sensitive lines: 0.32, *p=0.01, t-test).

Discovery of small molecules that induce sensitivity to chemotherapy

Patients whose tumors show resistance to front-line chemotherapy are generally given second- or third line therapy or recruited into clinical trials for experimental drugs (Broxterman and Georgopapadakou, 2005; Minna *et al.*, 2004). Furthermore, many patients who initially respond to front-line therapy suffer relapse, highlighting that additional therapeutic options are required. To identify additional therapeutics that might increase response to chemotherapy, we sought to discover small molecule candidates that might perturb the gene expression of chemotherapy resistant cell populations to induce sensitivity to chemotherapy. To this end we used the Connectivity Map (Lamb *et al.*, 2006) to search for compounds that reduce expression of resistance associated probe sets and increase the expression of sensitivity associated probe sets, using both our patient-derived and cell line-derived predictors of chemotherapy response. Notably, the latter analysis returned one compound that consistently showed high connectivity to

chemotherapy sensitivity, irrespective of the gene expression predictor that was used to probe the Connectivity Map (Fig. 3A).

We hypothesized that small molecule sensitizers to chemotherapy would independently have little effect on cell viability, but rather induce sensitivity to chemotherapy, manifested as a significant decrease in cell viability when used in combination with chemotherapy over that seen from chemotherapy alone. To learn whether any of the identified compounds had these characteristics, we measured the dose response of the HCC1954 breast tumor cell line to adiphenine using AlamarBlue proliferation assays (Fig. 3B). This data suggested that adiphenine is a relatively non-toxic small molecule that has little effect on cell viability at doses as high as 250μM. To learn whether adiphenine sensitized breast tumor cells we measured its capacity to potentiate doxorubicin induced cytotoxicity in HCC1954 and BT549 breast tumor cells, both of which we previously found to be relatively resistant to doxorubicin. Notably, adiphenine (250μM) had a negligible effect on HCC1954/BT549 cell viability, but dramatically increased the cell killing effect of doxorubicin (100nM) in both cell lines tested (Fig. 3C, HCC1954: *$p=0.0035$, BT549: *$p<0.0001$, t-test). Since, adiphenine is a reported inhibitor of nicotinic and muscarinic acetylcholine receptors (nAChR/mAChR), we tested the capacity of other structurally related and structurally distinct nAChR/mAChR antagonists to sensitize breast tumor cells to doxorubicin (Fig. 3 D). The IC50 of proadifen, a structural analog of adiphenine, was 45μM in HCC1954 cells, and we selected 25μM as the highest concentration of proadifen that had negligible effect on cell viability, to test for chemo-sensitizing capacity (Fig. 3E). Notably, in sensitization

assays, 25μM proadifen had little effect on cell viability, but like adiphenine, potentiated doxorubicin (100nM) cytoxicity in HCC1954 and BT549 breast tumor cells (Fig. 3F, HCC1954: *p=0.0004, BT549: *p<0.0001, t-test). To learn whether adiphenine and proadifen functioned through either nAChRs or mAChRs, we tested the capacity of NDNI, a structurally unrelated and specific inhibitor of nAChRs(Sumithran *et al.*, 2005), and we selected 6μM as the highest concentration of NDNI that had negligible effect on cell viability, to test for chemo-sensitizing capacity (Fig. 3G). As seen with adiphenine and proadifen, 6μM NDNI potentiated doxorubicin (100nM) cytoxicity in HCC1954 and BT549 breast tumor cells (Fig. 3H, HCC1954: *p = 0.01, BT549: *p = 0.003, t-test). Taken together, these data suggest that adiphenine, proadifen and NDNI function as chemotherapy sensitizers in human breast tumor cells. Furthermore, because NDNI is a reportedly specific inhibitor of nAChRs, chemosensitizing likely occurs through inhibition of nAChRs, rather than through mAChR inhibition.

To further confirm that chemo-sensitization by these compounds was mediated by its interaction with nAChRs as opposed to mAChRs and to gain insight into the nature of the relevant receptor, we surveyed the capacity of a greater number of muscarinic (mAChRs) and nAChR antagonists to sensitize HCC1954 breast tumor cells to doxorubicin. Whereas only one of the mAChR antagonists functioned as a chemo-sensitizer, many of the nAChR antagonists did (Table 1). Taken together with our previous data, these additional findings confirm that chemo-sensitization is mediated through specific inhibition of nAChRs.

Adiphenine and analogs target human breast tumor initiating cells

Given the role of breast TICs in mediating chemotherapy resistance, we sought to examine whether adiphenine might also target these cells. To this end, we downloaded the gene expression profiles of multiple breast tumors (n=14) sorted into TIC ($CD44^+/CD24^-$) and non-TIC (non- $CD44^+/CD24^-$) cell fractions (GSE7513) (Creighton *et al.*, 2009) (Fig. 4A). We hypothesized that if adiphenine targets TICs, it would likely reduce the expression of TIC specific transcripts and increase the expression of non-TIC transcripts. We identified differentially expressed genes between the TIC and non-TIC tumor fractions (Fig. 4B), and tested whether adiphenine showed transcriptional connectivity to these genes. Notably, adiphenine had a high positive connectivity score in the MCF7 breast tumor cell line (Mean: 0.82, *p = 0.0003), demonstrating that adiphenine treatment is associated with reduced TIC-specific gene expression, and increased non-TIC-specific gene expression, supporting the notion that adiphenine might target breast TICs. To test this possibility further, we completed sphere forming assays with adiphenine, proadifen and NDNI in multiple human breast tumor cell lines. Whereas the identity of cells that form spheres is not known with certainty, mammosphere and tumorsphere forming cells co-fractionate with mammary epithelial stem/progenitor cells, as well as breast TIC, respectively. Moreover, agents affecting the frequency of sphere-forming cells correspondingly alter the frequency of mammary epithelial stem/progenitor cells, as well as breast TIC, as measured by cell transplantation experiments, suggesting that sphere formation is a reasonable surrogate assay for measuring TICs.

Adiphenine, proadifen and NDNI inhibited sphere formation by cells from all cell lines tested (Fig. 4 D-F). Interestingly, the potency of these compounds for inhibition of sphere formation recapitulated that observed in the chemosentization assays. In both cases the dose required for either chemosensitization to doxorubicin or inhibition of sphere formation by adiphenine was substantially higher than those required for proadifen or NDNI. Moreover, HCC1954 spheres treated for 96 hours with proadifen did not form tumors when transplanted into mice, suggesting that proadifen targeted functional breast TICs in this cell line (Fig. 5A). Importantly, these data are consistent with the notion that adiphenine and adiphenine analogs target breast TICs, thereby accounting for their capacity to overcome chemotherapy resistance in human breast tumor cell lines.

Discussion

Cancer medicine faces the continual challenge of improving response rates to anti-cancer therapy. The first efforts of systemic treatment development gave rise to single agent chemotherapeutics such as cyclophosphamide (and other akylating agents,) as well as the anti-folates (aminopterin and methotrexate) which showed some efficacy in treating various tumor types (Chabner and Roberts, 2005; Farber and Diamond, 1948; Gilman, 1963). Whereas these agents generally caused temporary remission, many tumors subsequently progressed suggesting that they were unable to provide durable cure when used as single agents. This observation led to the development of combination chemotherapy. However, this strategy is still not always successful, especially in the context of advanced disease (Breathnach *et al.*, 2001; Herbst *et al.*, 2006). Some recent efforts have focused on profiling the molecular features of tumors, and identifying biomarkers that predict sensitivity to commonly used chemotherapeutic regimens (Ayers *et al.*, 2004; Hess *et al.*, 2006; Potti and Nevins, 2008; Staunton *et al.*, 2001). Other efforts have sought to identify compounds that sensitize tumors to chemotherapy without damaging normal tissue (Comen and Robson, 2010; Liang and Tan, 2010; Lord and Ashworth, 2008; Thangasamy *et al.*, 2008; Thangasamy *et al.*, 2010). Additional studies have focused on identifying agents that specifically target breast TICs (Gupta *et al.*, 2009), in order to provide durable cancer cure. The results presented here, focusing on improving treatment response to chemotherapy, combine the goals of all these efforts, and aimed to use predictive gene signatures and Connectivity Mapping as a way to identify compounds with that likely sensitize tumor cells to chemotherapy (Lamb *et al.*,

2006; Lehar *et al.*, 2007). Of note, this strategy resulted in the identification of adiphenine, which sensitized breast tumor cells to chemotherapy. Moreover, additional transcriptional analyses revealed that adiphenine also affected the expression of breast TIC specific transcripts, and ultimately targeted breast TICs *in vitro*. .

Importantly, adiphenine is reportedly an inhibitor of nAChR and mAChRs (Spitzmaul *et al.*, 2009) and is an orally active smooth muscle relaxant used widely to treat urinary, biliary, and intestinal colic (Gaensler and McGowan, 1950). Our data suggests that the use of adiphenine may improve patient response to the chemotherapy, possibly through the eradication of chemotherapy resistant breast TICs. In this regard it is useful to note that adiphenine is known to cross the blood-brain barrier(Michelot *et al.*, 1986), suggesting it will also be useful in targeting breast tumor metastases to the brain. While the molecular target of adiphenine is not known with certainty, we screened multiple additional nAChR/mAChRs antagonists, both related and unrelated structurally to adiphenine, and found that many nAChR antagonists similarly functioned as chemosensitizing agents, suggesting that chemosensitization is achieved through nAChR inhibition. nAChRs, which comprise 16 genes in mammals whose protein products function as homo- or hetero-pentameric ligand-gated ion channels (Albuquerque *et al.*, 2009; Taly *et al.*, 2009). The nAChR family has been broadly divided into neuronal-type and muscle-type subunits; with the exception of the muscle subtype, which comprises subunits encoded by 4 different genes, all the other subtypes are hetero-pentamers comprising pairs of subunits or homo-pentamers. The α-9 and α-10 subunits form the most disparate nAChR subfamily, and share sequence and pharmacological similarities to

mAChRs, gamma-aminobutyric acid-A, and serotonin type 3 receptors (Rothlin *et al.*, 1999; Verbitsky *et al.*, 2000). Whereas all nAChRs are ligand-gated channels that permit transport of cations (Na^+ and K^+), the α-7 and α-9 or α-9/α-10 nAChRs selectively transport Ca^{++}. Interestingly, the α9/α10 nAchR subtype has been characterized to have mixed nAChR/mAChR pharmacological properties, suggesting that mAChR inhibitors also inhibit these nAChRs. Indeed, this data might explain our observation that dicyclomine, a mAChR antagonist, also sensitized breast tumor cells to chemotherapy. This would also indicate that the revelant receptors for chemosensitization or targeting breast TICs would be α-9 or α-9/α-10 nAChRs, athough this is not known with certainty.

Interestingly, an emerging literature engendered by studies of the role of passive smoking in oncogenesis reveals that nAChRs play roles in several malignancies especially lung (Schuller, 2009) and breast cancer (Lee *et al.*, 2010; Wu *et al.*, 2011). The prevailing model suggests that agonists such as nicotine acting through the nAChR pathway cause Ca^{++} influx thereby activating signaling pathways that stimulate tumor cell proliferation, inhibit apoptosis and induce the release of angiogenic factors, processes which sustain tumor growth(Schuller, 2009). Interestingly α-9 nAChR transcripts and protein are overexpressed by a factor of ~8-fold in 67% of 276 human breast tumor samples examined (Lee *et al.*, 2010). Knockdown of α-9 nAChR transcripts with shRNAs in the MDA-MB-231 breast tumor cell line inhibited the nicotine-dependent formation of colonies in soft-agar suspension and the growth of breast tumor xenografts in mice. Furthermore, a cDNA encoding the α-9 nAChR endowed non-tumorigenic MCF-10A cells with the capacity to form colonies in soft-agar suspension dependent on the presence

of nicotine and increased the nicotine-dependent growth of MCF-10A tumor xenografts. Additional reports suggest that nicotine increases the proportion of cancer stem cells in human breast tumor cell lines (Hirata *et al.*, 2010), and can also decrease the cytotoxicity of chemotherapeutic drugs, *in vitro (Zhang et al., 2009)* .

In total, this study provides a straight-forward approach to use genomic based predictors for compound discovery. Here we present the use of this approach to identify chemotherapy sensitizing molecules that in this case, also target breast TIC. Based on these data w suggest that inhibition of nAChRs may prove to be a promising approach for anti-cancer therapy. Moreover, we also believe that the approach taken here can be generalized to multiple disease types, ultimately providing a method for efficient drug discovery across many disease types.

Methods

Cell viability screens: Human breast tumor cell lines were purchased from the ATCC and passaged according to their protocols. All screening was completed using a Biomek3000 (Beckman Coulter) and a DTX880 plate reader (Beckman Coulter). When cells had reached 80-90% confluency, they were trypsinized, washed and then seeded at 50,000 cells/ml into the wells of an opaque 384-well plate (Nunc) (50µl/well) in quadruplicate. Cells were allowed to adhere for 4 hours, followed by compound addition (50X concentration, 1µl/well). Parent stocks of A and T were dissolved from powder (Sigma) into DMSO and were diluted to 50X assay stocks the day of the assay. Positive (DMSO only) and negative controls (medium only) were also included in the plate. After 24 hours, 5µl of alamarBlue was added to every well, and alamarBlue reduction was read (530nM excitation / 590nM emission) after an additional 24 hours. Raw intensity values were converted to residual activity using the following formula:

$$\%\, residual\ acitivty = \frac{Sample\ reading - Negative\ control}{Positive\ control - Negative\ control} * 100$$

The IC_{50} of cell viability for each cell line was plotted by fitting unconstrained non-linear regression using GraphPad Prism 5™.

Gene expression experiments: Cell lines representing extremes of resistance and sensitivity were used as a training set to generate a gene signature predicting AT resistance. Cell line expression profiles were downloaded from ArrayExpress (E-TABM-157) and normalised using RMA(Irizarry *et al.*, 2003). Training patient data was downloaded from GEO (GSE25055), and normalised using RMA. pCR and RD were used as the relevant training variables. In both cases feature selection was performed

using Bayesian probabilistic models(West *et al.*, 2001). Patient validation data was similarly download from GEO, and normalised using RMA (GSE25065, GSE21094). Differences in predicted responders and non-responders was analysed using receiver-operator-characteristic curves (ROC), odds ratios, and a fisher's exact test. Breast TIC data was downloaded from GEO (GSE7513) and normalised using RMA.

Connectivity Mapping: The probe set signatures were divided into resistance probes (high expression was related to resistance), and sensitive probes (high expression was related to sensitivity), and used to search the Connectivity Map (Broad Institute) for compounds, which reduced the expression of resistance probes and increased that of sensitivity probes.

Connectivity Map compound screen: Compounds identified by Connectivity Mapping were purchased from Sigma and screened in a similar fashion as described above. Parent and assay stocks were prepared as described above. Compounds were added to cells treated with or without AT and cell viability was measured using alamarBlue.

Sphere forming assays: Four days old sphere cultures were trypsinized for several minutes and then mechanically dissociated and plated at 30,000 cells/mL in SCM (low glucose DMEM: Ham's F-12 (3:1) supplemented with 1mg/ml fungizone, 1% penicillin/streptomycin, 2 μg/ml B-27, 10 ng/ml human bFGF, 20 ng/ml human EGF, 5 μg/ml Insulin, 4 ng/ml Heparin) in 96-well flat-bottom plates [Corning]. The inhibitor was solubilized in DMSO and added to the plates at indicated concentration. Equal volume of DMSO was added to the control wells. All dishes were incubated at 37°C and

5% CO_2. The number of spheres formed was counted using a light microscope 7 days later.

Transplant assays: Dissociated sphere cells were incubated in the presence and absence of proadifen for 96 hours. Spheres that arose over this period were dissociated and resuspended in 50% Matrigel [BD], 45% PBS and 5% FBS [Invitrogen], and equal numbers of viable cells were transplanted into NOD/SCID mice [University Health Network]. Mice were palpated weekly to monitor the formation of tumors.

References

Albuquerque EX, Pereira EF, Alkondon M, Rogers SW (2009). Mammalian nicotinic acetylcholine receptors: from structure to function. *Physiological reviews* **89:** 73-120.

Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K *et al* (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* **22:** 2284-93.

Bosl GJ, Gluckman R, Geller NL, Golbey RB, Whitmore WF, Jr., Herr H *et al* (1986). VAB-6: an effective chemotherapy regimen for patients with germ-cell tumors. *J Clin Oncol* **4:** 1493-9.

Breathnach OS, Freidlin B, Conley B, Green MR, Johnson DH, Gandara DR *et al* (2001). Twenty-two years of phase III trials for patients with advanced non-small-cell lung cancer: sobering results. *J Clin Oncol* **19:** 1734-42.

Broxterman HJ, Georgopapadakou NH (2005). Anticancer therapeutics: "Addictive" targets, multi-targeted drugs, new drug combinations. *Drug Resist Updat* **8:** 183-97.

Chabner BA, Roberts TG, Jr. (2005). Timeline: Chemotherapy and the war on cancer. *Nat Rev Cancer* **5:** 65-72.

Comen EA, Robson M (2010). Poly(ADP-ribose) polymerase inhibitors in triple-negative breast cancer. *Cancer J* **16:** 48-52.

Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A *et al* (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A* **106:** 13820-5.

Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J *et al* (2011). Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol* **29:** 1578-86.

Devita VT, Jr., Serpick AA, Carbone PP (1970). Combination chemotherapy in the treatment of advanced Hodgkin's disease. *Ann Intern Med* **73:** 881-95.

Farber S, Diamond LK (1948). Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. *N Engl J Med* **238:** 787-93.

Gaensler EA, McGowan JM (1950). Activity of some newer spasmolytic agents on the biliary tract of man; a comparative study of isopropyl-norepinephrine, adiphenine hydrochloride, methylamino-iso-octene and khellin. *J Pharmacol Exp Ther* **99:** 304-11.

Gilman A (1963). The initial clinical trial of nitrogen mustard. *Am J Surg* **105:** 574-8.

Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA *et al* (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138:** 645-59.

Hallett RM, Dvorkin, A., Gabardo, C.M., Hassell, J.A. (2010). An algorithm to discover gene signatures with predictive potential. *Journal of clinical and experimental cancer research*.

Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A *et al* (2011). A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* **305:** 1873-81.

Herbst RS, Bajorin DF, Bleiberg H, Blum D, Hao D, Johnson BE *et al* (2006). Clinical Cancer Advances 2005: major research advances in cancer treatment, prevention, and screening--a report from the American Society of Clinical Oncology. *J Clin Oncol* **24:** 190-205.

Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA *et al* (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24:** 4236-44.

Hirata N, Sekino Y, Kanda Y (2010). Nicotine increases cancer stem cell population in MCF-7 cells. *Biochem Biophys Res Commun* **403:** 138-43.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al* (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249-64.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ *et al* (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313:** 1929-35.

Lee CH, Huang CS, Chen CS, Tu SH, Wang YJ, Chang YJ *et al* (2010). Overexpression and activation of the alpha9-nicotinic receptor during tumorigenesis in human breast epithelial cells. *Journal of the National Cancer Institute* **102:** 1322-35.

Lee JK, Coutant C, Kim YC, Qi Y, Theodorescu D, Symmans WF *et al* Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer. *Clin Cancer Res* **16:** 711-8.

Lehar J, Zimmermann GR, Krueger AS, Molnar RA, Ledell JT, Heilbut AM *et al* (2007). Chemical combination effects predict connectivity in biological systems. *Mol Syst Biol* **3:** 80.

Liang H, Tan AR (2010). Iniparib, a PARP1 inhibitor for the potential treatment of cancer, including triple-negative breast cancer. *IDrugs* **13:** 646-56.

Lord CJ, Ashworth A (2008). Targeted therapy for cancer using PARP inhibitors. *Curr Opin Pharmacol* **8:** 363-9.

Michelot J, Moreau MF, Veyre A, Labarre P, Meyniel G (1986). Brain uptake of labelled adiphenine in rats. *Biopharm Drug Dispos* **7:** 197-206.

Minna JD, Gazdar AF, Sprang SR, Herz J (2004). Cancer. A bull's eye for targeted lung cancer therapy. *Science* **304:** 1458-61.

Minocha A, Long BH (1984). Inhibition of the DNA catenation activity of type II topoisomerase by VP16-213 and VM26. *Biochem Biophys Res Commun* **122:** 165-70.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T *et al* (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10:** 515-27.

Potti A, Nevins JR (2008). Utilization of genomic signatures to direct use of primary chemotherapy. *Curr Opin Genet Dev* **18:** 62-7.

Rothlin CV, Katz E, Verbitsky M, Elgoyhen AB (1999). The alpha9 nicotinic acetylcholine receptor shares pharmacological properties with type A gamma-aminobutyric acid, glycine, and type 3 serotonin receptors. *Molecular pharmacology* **55:** 248-54.

Schuller HM (2009). Is cancer triggered by altered signalling of nicotinic acetylcholine receptors? *Nature reviews. Cancer* **9:** 195-205.

Spitzmaul G, Gumilar F, Dilger JP, Bouzat C (2009). The local anaesthetics proadifen and adiphenine inhibit nicotinic receptors by different molecular mechanisms. *Br J Pharmacol* **157:** 804-17.

Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J *et al* (2001). Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* **98:** 10787-92.

Sumithran SP, Crooks PA, Xu R, Zhu J, Deaciuc AG, Wilkins LH *et al* (2005). Introduction of unsaturation into the N-n-alkyl chain of the nicotinic receptor antagonists, NONI and NDNI: effect on affinity and selectivity. *AAPS J* **7:** E201-17.

Taly A, Corringer PJ, Guedin D, Lestage P, Changeux JP (2009). Nicotinic receptors: allosteric transitions and therapeutic targets in the nervous system. *Nature reviews. Drug discovery* **8:** 733-50.

Thangasamy T, Sittadjody S, Limesand KH, Burd R (2008). Tyrosinase overexpression promotes ATM-dependent p53 phosphorylation by quercetin and sensitizes melanoma cells to dacarbazine. *Cell Oncol* **30:** 371-87.

Thangasamy T, Sittadjody S, Mitchell GC, Mendoza EE, Radhakrishnan VM, Limesand KH *et al* (2010). Quercetin abrogates chemoresistance in melanoma cells by modulating deltaNp73. *BMC Cancer* **10:** 282.

Verbitsky M, Rothlin CV, Katz E, Elgoyhen AB (2000). Mixed nicotinic-muscarinic properties of the alpha9 nicotinic cholinergic receptor. *Neuropharmacology* **39:** 2515-24.

Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L *et al* (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* **20:** 719-26.

West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R *et al* (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A* **98:** 11462-7.

Wu CH, Lee CH, Ho YS (2011). Nicotinic acetylcholine receptor-based blockade: applications of molecular targets for cancer therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17:** 3533-41.

Zhang J, Kamdar O, Le W, Rosen GD, Upadhyay D (2009). Nicotine induces resistance to chemotherapy by modulating mitochondrial signaling in lung cancer. *Am J Respir Cell Mol Biol* **40:** 135-46.

Table and Figure legends

Table 1. Chemo-sensitization capacities of nAChR and mAChR inhibitors

Figure 1. Identification and validation of a gene signature predicting breast cancer patient response to neoadjuvant chemotherapy. A) Experimental strategy for identifying probe sets predictive of neoadjuvant chemotherapy response. **B)** Identification of probe sets related to response shared among ER+ and ER- breast tumors. **C)** Expression of predictive probe sets in GSE25055 (training cohort) for patients predicted to respond and not respond (left to right, respectively). Black vertical lines along bottom of heatmap indicate response. **D)** Performance of predictive probe sets identified in GSE25055 cohort on training data (GSE25055, *$p<0.0001$, OR: 12.3) and two additional validation datasets (GSE25065[n=198]: *$p=0004$, OR: 4.8. GSE21094[n=278]: *$p<0.0001$, OR: 9.5).

Figure 2. Identification and validation of a gene signature that predicts breast tumor cell line response to chemotherapy. A) Cell viability dose-response curves for each cell line treated with doxorubicin. **B)** Residual activity (Cell viability) of tested cell lines at 100nM doxorubicin. **C)** Experimental strategy to identify probe sets associated with cell line resistance/sensitivity to doxorubicin. **D)** Gene expression of top 250 probe sets associated with cell line resistance/sensitivity to doxorubicin. **E)** Leave-one-out cross-validation of cell lines used to identify the aforementioned 250 probe sets (*$p=0.01$, t-test).

Figure 3. Identification and experimental validation of candidate sensitizing molecules to chemotherapy. A) Connectivity Mapping with patient and cell line derived chemotherapy response signatures both identify adiphenine as a candidate chemosensitizing agent. B**)** IC50 curve for adiphenine in HCC1954 breast tumor cells. **C)** 250μM adiphenine has negligible of cell viability but sensitizes cells to doxorubicin induced cytotoxicity in HCC1954 and BT549 cells (HCC1954: *p=0.004, BT549: *p<0.0001). **D)** Chemical structures of adiphenine, proadifen and NDNI. **E)** IC50 curved for adiphenine in HCC1954 breast tumor cells. **F)** 25μM proadifen has a negligible effect on cell viability but sensitizes cells to doxorubicin induced cytotoxicity in HCC1954 and BT549 cells (HCC1954: *p=0.0004, BT549: *p<0.0001). **G)** IC50 curve for NDNI in HCC1954 breast tumor cells. **H)** 6μM NDNI has a negligible effect on cell viability but sensitizes cells to doxorubicin induced cytotoxicity in HCC1954 and BT549 cells (HCC1954: *p=0.01, BT549: *p=0.003).

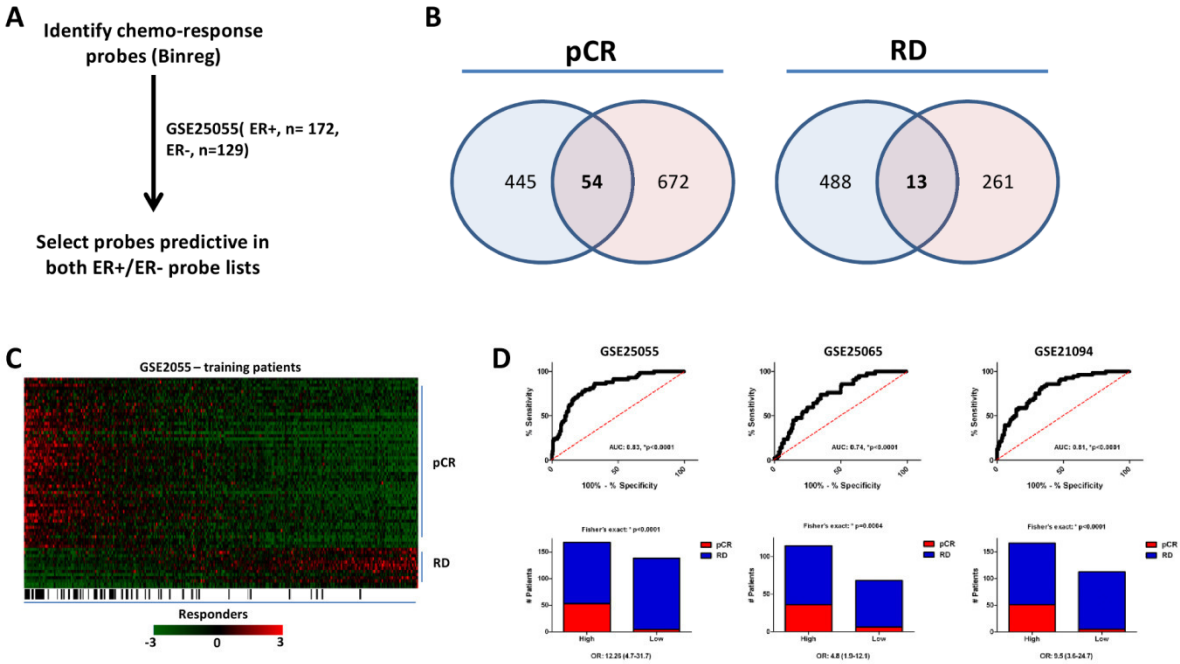
Figure 4. Adiphenine treatment is associated with reduced expression of breast TIC genes as well as inhibits sphere formation. A) Experimental strategy to learn whether adiphenine affects expression of breast TIC specific genes. **B)** Expression of TIC and non-TIC in 28 breast tumors. **C)** Connectivity map score of adiphenine in MCF7 cells to TIC/non-TIC genes (Mean: 0.82, *p=0.0003. **D)** Adiphenine inhibits sphere formation in multiple human breast tumor cell lines. **E)** Proadifen inhibits sphere formation in multiple human breast tumor cell lines. **F)** NDNI inhibits sphere formation in multiple human breast tumor cell lines.

Figure 5. Proadifen targets functional breast TICs in the HCC1954 human breast tumor cell lines. A) HCC1954 spheres treated with proadifen do not engraft and initiate tumor growth in NOD/SCID mice relative to vehicle treated counterpart HCC1954 cells

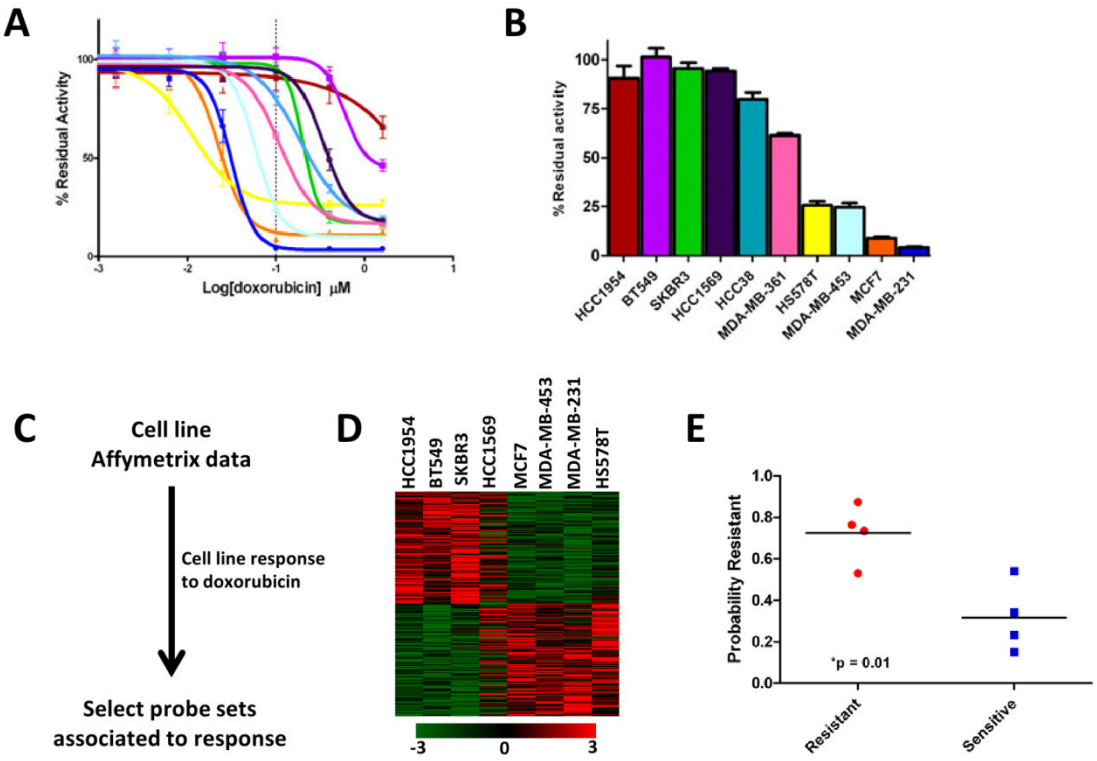
Table 1. Chemo-sensitization capacities of nAChR and mAChR inhibitors

Scopolamine	Non-selective muscarinic antagonist6	No
Ipratopium	Non-selective muscarinic antagonist7	No
Himbacine	M2/M4 muscarinic antagonist8	No
Atropine	Non-selective muscarinic antagonist9	No
Gallamine	Non-selective muscarinic antagonist10	No
Dicyclomine	Muscarinic Antagonists of the M1 receptor	Yes
Methyllycaconitine	α9/α7 nicotinic cholinergic receptor antagonist11	No
Bupropium	Highest affinity for α3β4 subtype of the nicotinic receptor12	No
Tubocurarine	α2 subunit of nicotinic receptor13	No
Mecamylamine	Partial selectivity for α7 nicotinic receptor14	No
Lobeline	Highest affinity for the α4β2/ α3β2 nicotinic receptor subunits15	Yes
Dihydro-β-Erythroidine	Highest affinity for the α4β4 and α4β2 nicotinic receptor subunits16	Yes
NDNI	Highest affinity for the α4β2 nicotinic receptor, compared to other nicotine analogs17	Yes

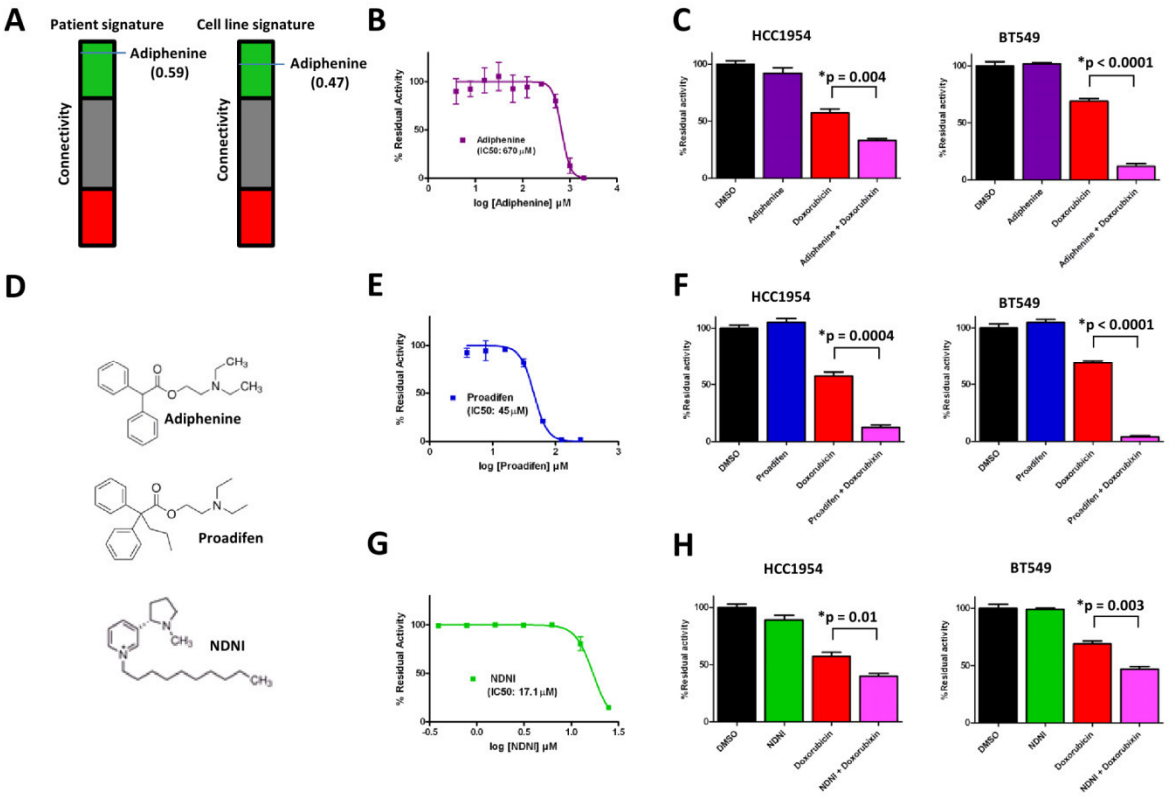
Hallett, *et al.* Figure 1



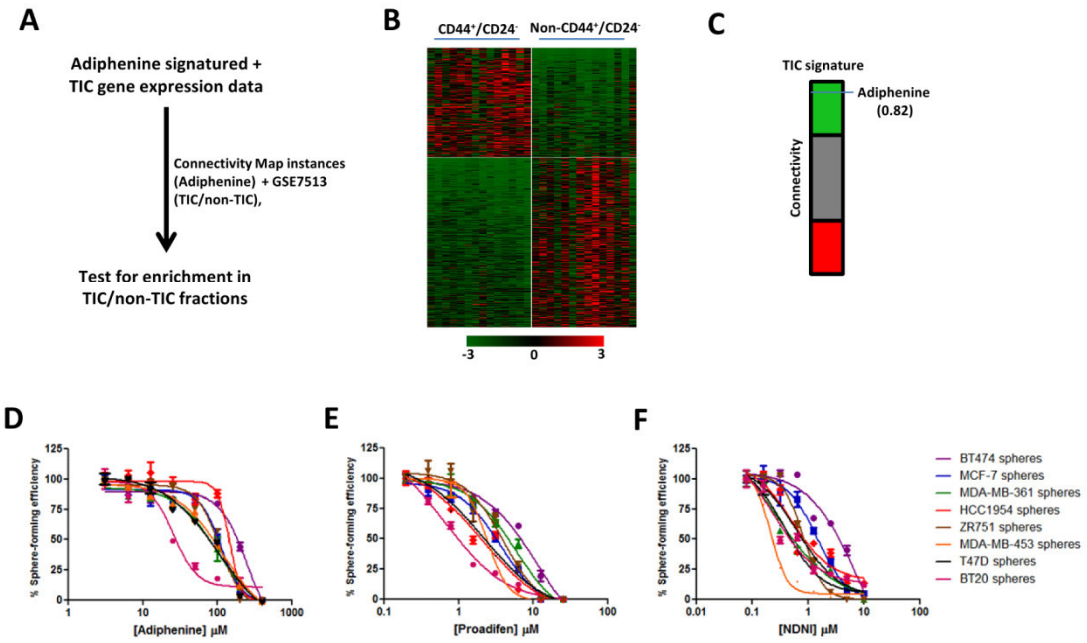
Hallett, *et al.* Figure 2



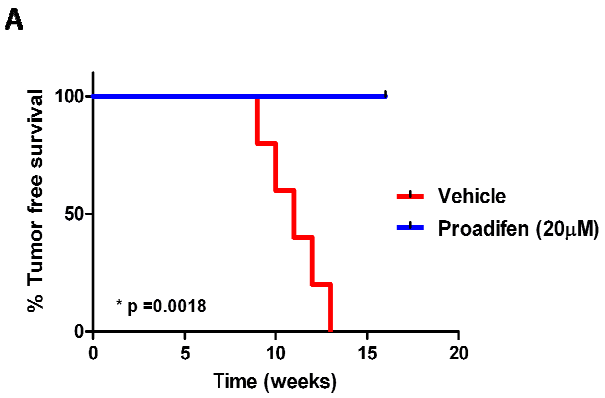
Hallett, *et al.* Figure 3



Hallett, *et al.* Figure 4



Hallett, *et al.* Figure 5



CHAPTER 6

Introduction

As described in the chapter 1, recent data suggests that tumor relapse is driven by a minority subpopulation of tumor cells termed cancer stem cells, or tumor-initiating cells (TIC), which resist currently available anti-cancer therapeutics, and therefore cause recurrence. Hence new therapies are needed that target TIC as well as the bulk tumor cells to provide durable breast cancer remission. Whereas the study of human breast TICs has been hampered by our inability to fractionate these cells into pure or nearly pure cell populations (Al-Hajj *et al.*, 2003; Battula *et al.*, 2012; Ginestier *et al.*, 2007), we've found that breast cancer-prone transgenic mice such as those that model ERBB2-positive breast cancer produce mouse mammary tumors rich in breast TICs (Kondratyev *et al.*, 2011). We found that mammary tumors from these transgenic mice (MMTV-*Neu)* comprise ~30% breast TICs and that cells from these tumors can be propagated *in vitro* as non-adherent tumorspheres, which also comprise a similarly high fraction of breast TICs (Kondratyev *et al.*, 2011). We have also used these same culture conditions to propagate mouse mammary epithelial stem and progenitor cell cultures as non-adherent mammospheres, which serve as a normal stem/progenitor cell controls to compliment our studies of breast TICs (Kondratyev *et al.*, 2011).

Using MMTV-*neu* tumor derived mouse mammary tumorspheres and strain matched mammospheres as *in vitro* models of enriched cultures of breast TICs and their normal mammary epithelial stem/progenitor cell counterpart, we identified that the

Wnt/β-catenin signaling pathway was hypertactive in breast TICs. Interestingly, there is significant data that Wnt/β-catenin signaling is involved in the maintenance of cancer stem cells (de Sousa *et al.*, 2011; Dodge and Lum, 2011; Korkaya *et al.*, 2009; Vermeulen *et al.*, 2010), and many groups are trying to develop Wnt/β-catenin signaling inhibitors for cancer treatment (Chen *et al.*, 2009; Eguchi *et al.*, 2005; Lepourcelet *et al.*, 2004). Many studies suggest that Wnt/β-catenin signaling plays a role in human breast tumors, as cytoplasmic localization of β-catenin, epigenetic silencing of pathway antagonists and LRP6 co-receptor expression are all linked to poor outcome (Liu *et al.*; Lopez-Knowles *et al.*; Suzuki *et al.*, 2008). Furthermore, inhibition of LRP6 inhibits tumor growth in xenograft models (Liu *et al.*), and active Wnt signalling has been implicated in mediating resistance to radiation therapy (Woodward *et al.*, 2007).

In this manuscript, we report that inhibitors of Wnt/β-catenin signaling selectively target mouse mammary tumor-derived TICs. Moreover, we report that one such inhibitor, PKF118-310, halts tumor growth *in vivo* and treated tumors are relatively devoid of functional TICs.

Manuscript #5

Title: Small molecule antagonists of the Wnt/B-catenin signaling pathway target breast tumor-initiating cells in a Her2/Neu mouse model of breast cancer

Authors: Robin M. Hallett[1], Maria K. Kondratyev[1], Andrew O. Giacomelli[1], Allison M. L. Nixon[1], Adele Girgis- Gabardo[1], Dora Ilieva[1], and John A. Hassell[1]*

1. Department of Biochemistry and Biomedical Sciences, Centre for Functional Genomics, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada L8N 3Z5

*Correspondence and requests for materials should be addressed to J.A.H. (hassell@mcmaster.ca)

.

Acknowledgements

This work was generously supported by grants from the Canadian Stem Cell Network, the Ontario Institute for Cancer Research and the Canadian Breast Cancer Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Robin Hallett was the recipient of a Canadian Breast Cancer Foundation Master's fellowship.

Author contributions

Robin Hallett - Conceived, planned, analyzed and performed most of the experiments in the paper, wrote the manuscript

Maria Kondratyev – Helped transplant cells for figure 3

Andrew Giacomelli – Obtained inhibitors from Novartis used throughout experiments

Allison Nixon – Isolated tumor cells for transplant used in figure 5

Adele Girgis-Gabardo – Helped with cell isolations and cell culture for cells used in figure 1 & 2

Dora Ilieva – Performed hybridizations and pre-processing of 'in house' microarray data used in figure 1

John Hassell – Conceived of and provided critical feedback during project, helped write manuscript

Abstract:

Background*:* Recent evidence suggests that human breast cancer is sustained by a minor subpopulation of breast tumor-initiating cells (BTIC), which confer resistance to anticancer therapies and consequently must be eradicated to achieve durable breast cancer cure.

Methods/Findings*:* To identify signaling pathways that might be targeted to eliminate BTIC, while sparing their normal stem and progenitor cell counterparts, we performed global gene expression profiling of BTIC- and mammary epithelial stem/progenitor cell-enriched cultures derived from mouse mammary tumors and mammary glands, respectively. Such analyses suggested a role for the Wnt/β-catenin signaling pathway in maintaining the viability and or sustaining the self-renewal of BTICs *in vitro*. To determine whether the Wnt/β-catenin pathway played a role in BTIC processes we employed a chemical genomics approach. We found that pharmacological inhibitors of Wnt/β-catenin signaling inhibited sphere- and colony-formation by primary breast tumor cells and primary mammary epithelial cells, as well as by tumorsphere- and mammosphere-derived cells. Serial assays of self-renewal *in vitro* revealed that the Wnt/β-catenin signaling inhibitor PKF118-310 irreversibly affected BTIC, whereas it functioned reversibly to suspend the self-renewal of mammary epithelial stem/progenitor cells. Incubation of primary tumor cells *in vitro* with PKF118-310 eliminated their capacity to subsequently seed tumor growth after transplant into syngeneic mice. Administration of PKF118-310 to tumor-bearing mice halted tumor growth *in vivo*.

Moreover, viable tumor cells harvested from PKF118-310 treated mice were unable to seed the growth of secondary tumors after transplant.

Conclusions*:* These studies demonstrate that inhibitors of Wnt/β-catenin signaling eradicated BTIC *in vitro* and *in vivo* and provide a compelling rationale for developing such antagonists for breast cancer therapy.

Results

Expression of Wnt/β-catenin signaling pathway components and target genes

Numerous previous studies have linked activation of Wnt/β-catenin signaling with breast cancer (Brennan and Brown, 2004; Clevers, 2006; Howe and Brown, 2004; Khramtsov *et al.*; Klarmann *et al.*, 2008; Lindvall *et al.*, 2006; Liu *et al.*, 2004; Nusse, 2005; Suzuki *et al.*, 2008; Ugolini *et al.*, 2001; Wong *et al.*, 1994; Woodward *et al.*, 2007). We sought to extend these findings by first investigating the expression of Wnt/β-catenin signaling pathway components in mouse tumorspheres, mammospheres, and mammospheres induced to differentiate *in vitro*, which we used as approximate *in vitro* models of BTICs, mammary epithelial stem and progenitor cells, and differentiated mammary epithelial cells, respectively. A microarray analysis of 3 independent tumorsphere cultures established from independent mouse mammary tumors and 3 separate preparations of mammospheres and mammospheres induced to differentiate *in vitro*, revealed that the transcripts of many Wnt/β-catenin signaling pathway genes were most highly expressed in tumorspheres relative to either mammospheres or mammospheres induced to differentiate (Fig. 1A). The latter included upstream pathway components such as Wnt ligands, receptors, as well as the transcriptional co-activator TCF4 (Fig. 1A). Several Wnt/β-catenin target genes were also most highly expressed in tumorspheres, including Axin2, cyclinD1 and CD44. Interestingly, the expression of transcripts encoding inhibitory Wnt/β-catenin signaling pathway elements, such as Sfrp1,

Srfp2 and Dkk2, were highest in mammospheres induced to differentiate, compared to both mammospheres and tumorspheres (Fig. 1A).

To verify the global gene expression profiling data we performed quantitative RT-PCR with primers that identified transcripts encoding components and downstream targets of the Wnt/β-catenin signaling pathway using additional independent preparations of RNA isolated from tumorspheres (n=3), mammospheres (n=3), and mammospheres induced to differentiate (n=3), respectively. These analysis confirmed our original findings, namely that transcripts of components and target genes of the Wnt/β–catenin signaling pathway were generally more highly expressed in tumorspheres compared to either mammospheres or mammospheres induced to differentiate *in vitro* (Fig. 1B-E, *$p<0.05$ for all genes shown, t-test). For example, Wnt7A and 7B, as well as the Fzd4 and 6 receptors were most highly expressed in tumorspheres compared to either mammospheres or mammospheres induced to differentiate *in vitro* (Fig. 1 B&D). By contrast, mammospheres and mammospheres induced to differentiate expressed higher levels of transcripts that encode negative regulators of the Wnt pathway, such as Frzb, Sfrp2 and Sfrp4 (Fig1. C & E).

We also investigated whether the finding of differential expression of Wnt/β-catenin signaling pathway components and target genes using the *in vitro* models were relevant to human breast cancer patients. To this end we used whole tumor gene expression profiles derived from human breast tumor RNA samples with accompanying overall- and metastasis free survival data (Chang *et al.*, 2005). We identified a Wnt-based

BTIC gene signature that comprised differentially expressed genes between tumorspheres and mammospheres that were induced to differentiate *in vitro.* We mapped these genes onto their human orthologs present in the previously published NKI whole tumor gene expression data set (Table 1). We then used the gene signature to learn whether we could separate patient tumor specimens from the NKI data set into tumorsphere-related and mammospheres induced to differentiate-related groups (tumorsphere-unrelated) (Hallett *et al.*, 2010). Notably, the expression of tumorsphere-related Wnt/β-catenin signaling pathway genes was linked to poor overall survival (Fig. 1F, *p=0.0005, log-rank test) and decreased metastasis-free survival (Fig. 1G, * p = 0.0068, log-rank test) when compared with the expression of mammospheres induced to differentiate-related Wnt/β-catenin signaling pathway genes. These data suggest that activated Wnt/β-catenin signaling may be a unique feature of BTICs derived from mouse mammary tumors relative to mammary epithelial stem and progenitor cells, and that Wnt/β-catenin pathway activation in BTICs may be linked to human breast cancer patient outcome and metastasis.

Wnt/β-catenin pathway agonists and antagonists regulate self-renewal and proliferation of BTIC and mammary epithelial stem/progenitor cells

To determine whether Wnt/β-catenin signaling was required for BTIC and mammary epithelial stem and progenitor cell function *in vitro*, we examined the effect of agonists and antagonists of Wnt/β-catenin signaling on sphere formation, which when performed under appropriate conditions (Pastrana *et al.*, 2011) is thought to be an attribute of stem/progenitor cells and TICs (Clevers; Hirschhaeuser *et al.*; Pastrana *et al.*,

2011; Reynolds and Rietze, 2005). We seeded dispersed cells dissociated from mammospheres and tumorspheres into media containing either Dkk1 or Wnt3a at concentrations previously shown to be sufficient to inhibit or activate the Wnt/β-catenin signaling pathway, respectively (Binnerts *et al.*, 2009; Reya *et al.*, 2003). Dkk1 is a secreted protein inhibitor (Glinka *et al.*, 1998), and Wnt3A is a stimulatory ligand of Wnt/β-catenin signaling. Addition of Dkk1 reduced sphere formation, whereas addition of Wnt3A stimulated sphere formation of both mammosphere- and tumorsphere-derived cells (Fig 2A). We used the small molecule BIO to inhibit GSKβ, a negative regulator of Wnt/β-catenin signaling (Meijer *et al.*, 2003). Addition of BIO at various concentrations stimulated sphere formation by both mammospheres- and tumorsphere-derived cells in a dose-dependent manner (Fig 2B).

We similarly tested pharmacological agents, which target Wnt/β-catenin signaling (PKF118-310, PKF115-584, CGP049090), for their affect on sphere formation. These inhibitors are specific antagonists of Wnt/β-catenin signaling that interrupt the penultimate step in pathway activation, namely the interaction between β-catenin and Tcf/Lef transcription factors (Lepourcelet *et al.*, 2004). We seeded freshly isolated primary tumor cells and primary mammary epithelial cells, as well as mammosphere- and tumorsphere-derived cells into medium containing various concentrations of the inhibitors. Each inhibitor reduced sphere formation by primary tumor cells and primary mammary epithelial cells (Fig. 2C), as well as by mammosphere- and tumorsphere-derived cells (Fig. 2D), in a dose dependent fashion. Notably, the inhibitory concentration of the compounds required to reduce sphere formation by 50% (IC_{50}) (Table 2) did not

significantly differ from those IC_{50} values reported previously to affect reduce Tcf-dependent reporter gene expression in cell lines (Table 2) (Lepourcelet *et al.*, 2004).

We also assessed the effects of the compounds on colony formation, an assay commonly used to enumerate stem and progenitor cells (Stingl, 2008). We seeded freshly isolated primary tumor cells and primary mammary epithelial cells, as well as mammosphere- and tumorsphere-derived cells at clonal cell density into collagen-coated plates in serum-containing medium. When colonies appeared, generally a week later, they were stained and counted. The inhibitors reduced colony formation by primary tumor cells and mammary epithelial cells (Fig. 2E), as well as that of mammosphere- and tumorsphere derived cells in a concentration dependent fashion (Fig. 2F). Inhibition of both colony and sphere formation occurred at similar inhibitor concentrations, suggesting that the inhibitors targeted both progenitor cells and stem cells as well as their tumor equivalent counterparts.

To learn whether the inhibitory compounds affected the self-renewal of sphere-forming cells, we investigated the capacity of cells exposed to the compounds to serially form spheres. In short, dispersed primary mammary epithelial cells and primary tumor cells were exposed to PKF118-310 during a 4-day primary sphere-forming assay. The spheres that formed were counted, the cells dissociated from the spheres and the dispersed cells plated to form secondary spheres in fresh medium lacking the inhibitors. As we previously demonstrated, PKF118-310 reduced primary sphere-formation by both primary tumor cells and primary mammary epithelial cells dependent on its concentration

(Fig. 3A&B). However, whereas the primary mammary epithelial cells treated with 1μM PKF118-310 formed new mammospheres in secondary sphere-forming assays at the same frequency as the vehicle-treated mammary epithelial cells (Fig. 3C), the primary tumor cells exposed to 1μM PKF118-310 exhibited a ~10 fold reduced capacity to form spheres relative to the vehicle-treated primary tumor cells (Fig. 3D, *$p < 0.05$). These observations demonstrate that a single exposure of tumor cells to PKF118-310 was sufficient to block their capacity to subsequently form spheres, even after PKF118-310 was no longer present in the medium. Conversely, mammary epithelial cells similarly exposed to PKF118-310 were not impaired in their sphere forming capacity, suggesting that the irreversible effect of PKF118-310 was limited to the primary tumor cells. These findings suggest that PKF118-310 reversibly affected the self-renewal of mammary epithelial cells *in vitro* but irreversibly affected the self-renewal of the primary breast tumor cells.

Our data suggesting that knock down of Wnt/β-catenin signaling by PKF118-310 irreversibly blocked the self-renewal of tumorsphere-initiating cells *in vitro,* prompted us to test whether this inhibitor specifically affected the capacity of BTICs to elicit tumor growth in syngeneic mice after tumor cell transplant. We incubated freshly isolated primary tumor cells under sphere forming conditions in presence of vehicle (DMSO) or PKF118-310 (1μM or 2.5 μM concentrations). After 4 days the spheres were collected, dissociated and equal numbers (5,000) of viable PKF118-310- and vehicle-treated cells were injected sub-cutaneously into the syngeneic mice. Upon endpoint, when the tumor in any individual mouse reached 10% of its weight (~6 weeks) all the animals were

sacrificed and the tumors harvested. Tumors that arose from tumor cells incubated with PKF118-310 (1μM) appeared with longer latency and were approximately 50% smaller at endpoint than those that arose from vehicle-treated tumor cells (control [m = 0.49 g], treatment [m = 0.26g], *p < 0.05) (Fig. 3E). Furthermore, when higher doses of PKF118-310 (2.5μM) were used in these experiments, the tumor cells exposed to this concentration of the compound failed to initiate tumor growth in any of the host mice (Fig. 3F, *p<0.05, ~6 weeks). Collectively, these data demonstrate that incubation of primary tumor cells with PKF118-310 for 4 days *in vitro* substantially reduced BTIC frequency.

PKF118-310 halts tumor growth *in vivo.*

Our data suggested that incubation of primary breast tumor cells *in vitro* with PKF118-310 substantially reduced BTIC frequency as assessed by cell transplant into syngeneic mice. To determine whether PKF118-310 affected the growth of pre-established breast tumors and reduced BTIC frequency *in vivo*, we transplanted primary breast tumor cells into syngeneic mice and treated these hosts with PKF118-310 after the tumors had reached a volume of ~1 cm^3. We found that treating tumor-bearing mice with 0.85mg/kg of PKF118-310 for 12 days (5 days on, 2 days off and 5 days on) inhibited tumor growth compared to their vehicle-treated counterparts, but did not induce significant tumor regression during the treatment period (Fig. 4A). After treatment ceased the tumors in mice administered the vehicle were ~3 fold larger than those in PKF118-310-treated mice (Fig. 4B, * p < 0.05)

To uncover potential mechanisms whereby the compound halted tumor growth, we prepared sections from the tumors of both cohorts and stained them with Hematoxylin and Eosin (H&E). Interestingly the tumors of the mice treated with PKF118-310 contained many cell-free areas that were often circumscribed by a ring of cells resembling a duct (Fig. 4C). In some cases the cell-free areas contained material, characteristic of milk protein secretions. Interestingly, the changes in histology induced by treatment with PKF118-310 are consistent with reduced tumor grade, a histological parameter associated with better survival outcomes in human breast cancer patients (Bloom and Richardson, 1957). We also examined whether inhibition of tumor growth was accomplished by a reduction in cell proliferation and/or induction of apoptosis. We stained tumor sections for markers of proliferation (Ki67) and apoptosis (cleaved caspase-3 and TUNEL). We observed a significant decrease in the frequency of Ki67 positive tumor cell nuclei in the PKF118-310 treated tumor-bearing mice compared to their vehicle-treated counterparts (Fig. 4D&E). We did not observe any positive staining for cleaved caspase-3 or TUNEL in tumors from either vehicle- or PKF118-310-treated mice (data not shown). Tumors from PKF118-310 treated mice comprised approximately 3-4-fold fewer Ki67-positive cells than tumors from vehicle-treated mice. We similarly simultaneously stained the tumor sections with antibodies to a luminal lineage marker (CK8) and with those to each of two myoepithelial lineage markers (CK14 and alpha-smooth muscle actin [alpha-SMA). The vast majority of the cells in sections prepared from tumors of the vehicle-treated mice expressed only the luminal lineage marker in keeping with previous findings (data not shown). Tumor sections prepared from tumors

of the mice administered PKF118-310 also only expressed the luminal lineage marker. Surprisingly the cells comprising the duct-like structures found in tumor sections of mice administered PKF118-310 expressed the luminal lineage marker but not either of the myoepithelial lineage markers.

To learn whether administration of PKF118-310 to tumor-bearing mice inhibited Wnt/β-catenin signaling in tumors, we measured the abundance of Wnt/β-catenin target gene transcripts in the tumor cells. The abundance of both axin2 and cyclin D1 transcripts was significantly lower in tumors harvested from PKF118-310-treated mice compared to their vehicle-treated counterparts, confirming that PKF118-310 targeted Wnt/β-catenin signaling *in vivo* (Fig. 4F).

The principal objective of our experiments was to learn whether inhibition of Wnt/β-catenin signaling targeted BTICs in tumors. Because, treatment with PKF118-310 did not completely shrink tumors *in vivo,* we wondered whether PKF118-310 eradicated functional BTICs in the tumors of treated mice. We reasoned that if PKF118-310 selectively targeted BTICs *in vivo,* viable cells isolated from PKF118-310 treated tumor-bearng hosts would engraft and elicit tumor growth less efficiently relative to their vehicle-treated counterparts. To this end, we transplanted tumor cells by injecting them sub-cutaneously (n=20, PKF118-310-treated; n=20, vehicle-treated) between the shoulders of syngeneic mice (10,000 cells/mouse) and measured tumor latency in the PKF118-310-treated and vehicle-treated transplant cohorts. Mice transplanted with tumor cells harvested from vehicle-treated mice experienced a median of 5-week tumor

free survival and all mice had palpable tumors after 9 weeks. By contrast, tumor cells harvested from PKF118-310 treated mice generally failed to initiate tumor growth. In fact, 85% of mice transplanted with PKF118-310 treated tumor cells remained tumor free over a 12 week follow-up period (Fig. 5). The tumors arising following cell transplant were invariably lodged in the #2 mammary fat pad, which extends from the ventral to the dorsal area of the host mice (data not shown).

Together with our previous observations, these results demonstrate that not only was PKF118-310 treatment sufficient to halt tumor-growth *in vivo,* but that cells comprising the tumors of mice administered the compound were substantially diminished in their capacity to engraft and initiate tumor growth compared to their vehicle-treated counterparts.

Discussion

Whereas TIC have been identified in a wide variety of human and mouse malignancies (Al-Hajj *et al.*, 2003; Cho *et al.*, 2008; Collins *et al.*, 2005; Fang *et al.*, 2005; Lapidot *et al.*, 1994; Liu *et al.*, 2007; O'Brien *et al.*, 2007; Singh *et al.*, 2003; Zhang *et al.*, 2008)**,** little is known about their underlying biology, and few compounds have been identified that selectively target these cells (Gupta *et al.*, 2009; Guzman *et al.*, 2005). Targeting TICs is an important cancer therapeutic objective as these cells are resistant to current cancer therapies, including chemo- and radiation-therapy (Bao *et al.*, 2006; Baumann *et al.*, 2008; Eramo *et al.*, 2006; Ma *et al.*, 2008). Hence whereas standard therapies result in tumor shrinkage, they may fail to provide long lasting cures because rare TIC survive and seed tumor relapse.

The use of genomic and drug discovery technologies, such as global gene expression profiling and high-throughput screening would greatly aid the search for anti TIC therapies. However, implementation of these methodologies has been confounded by a lack of suitable human BTIC-enriched populations for study. Typically BTIC represent an exceedingly small percentage (~0.01%) of the total tumor cell population (Al-Hajj *et al.*, 2003; Ginestier *et al.*, 2007), and even in the most highly enriched populations, BTIC rarely achieve more than 1-2% purity (Al-Hajj *et al.*, 2003; Ginestier *et al.*, 2007). Moreover, means of propagating BTIC-enriched tumor cell populations *in vitro* have not been described. To overcome these obstacles, we have studied BTICs from mouse mammary tumors of breast cancer prone transgenic models because they comprise a high

BTIC frequency, averaging ~30% in most tumors and companion tumorspheres (Kondratyev *et al.*, 2011).

To determine whether the Wnt/β-catenin pathway is required for the survival and/or self-renewal of BTIC, we employed three small-molecular weight tool compounds, PKF118-310, PKF115-584 and CGP049090, which were originally identified in a high throughput screen to identify those that abrogate the binding of β-catenin to Tcf4 *in vitro* (Lepourcelet *et al.*, 2004). Follow up analyses of these compounds revealed their capacity to: block β-catenin binding to GST-Tcf4 *in vivo*; reduce expression of a Wnt/β-catenin luciferase reporter; restore the β-catenin induced axis duplication of *Xenopus* embryos when co-injected with β-catenin; inhibit expression of the Wnt target genes *Myc* and *CyclinD1*; and retard the proliferation of colon cancer cell lines known to display hyperactive Wnt signaling *in vitro* (Lepourcelet *et al.*, 2004). Collectively, the latter findings suggest that PKF118-310, PKF115-584 and CGP049090 reduce Wnt/β-catenin signaling leading to the inhibition of cancer cell line proliferation *in vitro*. To the best of our knowledge the effect of these compounds on breast tumorigenesis has not previously been assessed.

Several studies have implicated Wnt/β-catenin signaling in both the pathogenesis of breast cancer and the regulation of normal mammary epithelial stem cell processes (Brennan and Brown, 2004; Lindvall *et al.*, 2007; Nusse, 2005). Our data suggests that the Wnt/β-catenin pathway is hyperactive in BTIC compared to normal mammary epithelial stem/progenitor cells or to their more differentiated descendants. We made use

of the small molecule inhibitors to investigate the consequences of inhibiting Wnt/β-catenin signaling in both breast tumor cells and normal mammary epithelial stem/progenitor cell populations. Due to the limited availability of the natural compounds PKF115-584 and CGP049090, we focused primarily on the use of PKF118-310, which can be chemically synthesized. Our initial experiments showed that each of the 3 compounds inhibited sphere and colony formation by primary tumor cells and primary mammary epithelial cells, as well as by established tumorsphere- and mammosphere-derived cells without any apparent selectivity. However, both PKF115-584 and CGP049090 displayed somewhat increased selectivity of between 6-7 fold ($IC_{50}^{MMS}/IC_{50}^{TMS}$, Table 1) for primary tumor cells over primary mammary epithelial cells in sphere forming assays compared to PKF118-310 (2-3 fold selectivity), indicating that further investigation of the potential selectivity these compounds is warranted.

We did not observe any significant selectivity of PKF118-310 for either the survival and/or self-renewal of tumorsphere-initiating cells compared to mammosphere-initiating cells in primary sphere-forming assays (Fig. 2C-F). However, a single exposure of primary tumor cells to PKF118-310 in a primary sphere-forming assay was sufficient to block subsequent secondary sphere formation in the absence of the compound. By contrast, mammary epithelial cells exposed to PKF118-310 were not impaired in their capacity to form secondary spheres, suggesting that the effect of PKF118-310 on secondary sphere formation is specific to BTICs. Taken at face value these observations suggest that PKF118-310 inhibited tumorsphere formation by an irreversible mechanism, whereas the compound acted reversibly to affect mammosphere formation. Inhibition of

Wnt/B-catenin signaling by PKF118-310 may be cytotoxic for tumorsphere-initiating cells, perhaps because they are addicted this pathway, whereas pathway inhibition may be cytostatic for mammosphere-initiating cells.

Whereas sphere formation is a convenient and relatively rapid surrogate *in vitro* assay for stem/progenitor and TIC activity, the nature of sphere-forming cells is controversial and consequently we employed additional means to identity the tumor cells that might be targeted by PKF118-310 (Clevers; Hirschhaeuser *et al.*; Pastrana *et al.*, 2011; Reynolds and Rietze, 2005; Reynolds and Weiss, 1992). To this end we transplanted primary tumor cell populations that had been incubated with PKF118-310 under the same conditions as had been used in primary sphere-forming assays and thereafter measured the capacity of the remaining viable tumor cells to seed tumor growth after transplantation into syngeneic FVB/N female mice. These transplantation assays directly assess BTIC frequency and demonstrated that PKF118-310 targeted these cells as manifested by a concentration-dependent reduction in tumor incidence in recipient mice resulting from transplant of the compound-treated tumor cells.

Administration of PKF118-310 to tumor-bearing mice blocked tumor growth during the 10-day treatment period, an interval during which the tumors expanded by 2-3 fold in tumor-bearing mice that were administered the vehicle. Histological analyses of tumor sections from mice administered PKF118-310 revealed loss of tumor architecture manifested as reduced cellularity and phenotypic features associated with reduced tumor grade. Whereas no evidence of apoptotic cell death or altered expression of

differentiation markers was evident in tumor sections from mice administered the compound (data not shown), the frequency of Ki67-positive cells, a biomarker of cell proliferation, was markedly reduced. Importantly, tumor cells harvested from mice exposed to PKF118-300 formed tumor grafts in only 3 of 20 mice transplanted with these cells some 2 weeks after tumors had already formed in 20/20 mice transplanted with vehicle-treated tumor cells. Our transplantation assay can detect single tumor cells in the bulk tumor cell population, which on average comprise ~30% functional BTIC as established by limiting dilution cell transplantation assays (Kondratyev *et al.*, 2011). Hence, our findings suggest that the frequency of BTIC comprising the tumors was dramatically reduced by inhibiting Wnt/B-catenin signaling in tumors (Kondratyev *et al.*, 2011). Taken together these multiple lines of investigation suggest that antagonists of Wnt/B-catenin signaling target BTIC and provide proof-of-principle that eradicating these cells leads to durable breast cancer remission.

Methods

Care and treatment of animals: All mice used in these experiments were housed in a Canadian Council on Animal Care (CCAC)-approved facility at McMaster University. Mice were provided with food and water *ad libitum*. All animal experiments were conducted in accordance with the requirements of the CCAC.

Tumor and mammary epithelial cell culture: The #3 and #4 mammary glands from virgin female FVB/N mice (6-8 weeks old) and mammary tumors were isolated as described previously (Kondratyev *et al.*, 2011; Pullan, 1996). Mammospheres and tumorspheres were established from the bulk primary mammary epithelial and tumor cell population respectively as described previously (Kurpios *et al.*, 2009). Serial passage of the mammospheres and tumorspheres was accomplished by mechanically dissociated the cells from spheres using titruation and reseeding the dispersed cells into fresh medium. Passage of the spheres was limited to 3-5 serial passages before the cells were harvested and RNA prepared. To induce a differentiation program in mammospheres *in vitro,* intact mammospheres were collected by centrifugation, the spheres were dissociated and the dispersed cells were plated at a density of 150,000 cells/ml on rat-tail collagen (Roche, Basel, Switzerland) coated 60mm Petri dishes (Dontu *et al.*, 2003; Stingl *et al.*, 1998). The cells were incubated for a week before they were harvested and used to prepare cellular RNA for analyses.

RNA isolation and analyses: Total RNA was isolated from tumorspheres, mammospheres or mammospheres induced to differentiate using an RNAeasy mini prep

kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. RNA was quantified using spectrophotometic analyses (A_{260nm}/A_{280nm}) and it's quality assessed by gel electrophoresis. RNA from 3 independent tumorsphere, mammosphere, and mammospheres induced to differentiate cell populations was used to prepare cRNA probes for hybridization to MOE430A Gene Chips™ in accordance with manufacturer's protocols (Affymetrix, Santa Clara, California). Gene expression profiling data was analyzed using Genespring™. Gene expression values were normalized to the average expression of either mammospheres induced to differentiate or tumorspheres for each probe set to generate a heat map. When a gene was represented by multiple probe sets, the most highly differentially expressed probe set was chosen for display in the heat map. These data have been deposited in the gene expression omnibus (GEO, GSE32463). Independent RNA preparations from different populations of tumorspheres, mammospheres, or mammospheres induced to differentiate was also analyzed by quantitative RT-PCR using the mouse Wnt Signaling Pathway RT^2 *Profiler*™ (QIAGEN). Total cellular RNA was isolated using the RNAeasy mini prep kit (QIAGEN) and used as template for oligo-dT primed reverse transcription using SuperScriptII™ First Strand Synthesis (Invitrogen, Carlsbad, California) for quantitative RT-PCR The abundance of selected mRNA transcripts were determined (primer sequences available upon request) with quantitative RT-PCR using FastStart DNA Master SYBR Green I Kit on the Light Cycler (Roche) according to the manufacturer's protocol.

Gene signature

Microarray and clinical data was downloaded from http://microarray-pubs.stanford.edu/wound_NKI/explore.html. The expression of Wnt/β-catenin signaling pathway genes was used to divide patients into related and unrelated Wnt/β-catenin signature groups as previously described (Hallett *et al.*, 2010).

Sphere, colony, and *ex vivo* treatment assays

Sphere and colony forming assays were completed as previously described (Kondratyev *et al.*, 2011). Dkk1 and Wnt3A were obtained from RnDSystems (Minneapolis, Minnesota). PKF118-310, PKF115-584 and CGP049090 were a gift from Novartis (Basel, Switzerland).

IC_{50} calculations

The 50% inhibitory concentration (IC_{50}) of compounds was calculated using GraphPad Prism5 software. X-axis values were X=Log(X) transformed and then fit with a dose-response curve. The DMSO vehicle control was included to aid IC_{50} calculation and was assigned a 1nM concentration of the tested compound.

***In vivo* compound administration**

Freshly isolated primary tumor (100,000) cells were suspended in 50% Matrigel (BD, Franklin Lakes, New Jersey), 45% phosphate buffered saline pH7.4 (PBS) and 5% fetal bovine serum (FBS) [Invitrogen, Carlsbad, California], and the cells were injected subcutaneously between the shoulders into syngeneic 6-8 week-old female mice (FVB/N

strain). Mice were monitored by palpation weekly for the occurrence of tumors. When tumors reached roughly $1cm^3$, the mice were administered either the vehicle (0.1% DMSO) or PKF118-310 (0.85mg/kg) dissolved in 0.1% DMSO by intra-tumoral injection for 5 consecutive days followed by a 2-day rest period before compound administration was repeated once. Tumor volume was measured twice weekly. At the end of the 12-day treatment cycle the mice were sacrificed and their tumors harvested for analysis. We found that tumors were invariably embedded in the fat pads of the number 2 mammary glands. Viable tumor cells (assessed by TrypanBlue staining) from vehicle- and compound-treated mice were isolated as described above and 10,000 cells were transplanted subcutaneously into syngeneic mice (n=20, treated; n=20, untreated).

Histology and immuno-histochemical analysis

Paraformaldehyde fixed tumor fragments were embedded in paraffin, sectioned and stained with H&E. The tumor sections were de-paraffinized and rehydrated in ethanol (100-70% gradient) before immunofluorescent analysis. Antigen retrieval was performed in Antigen Unmasking Solution (Vector, Burlingame, California). Slides were blocked with 3% normal goat serum (Dako, Denmark) and incubated with primary antibodies for 2 hours at room temperature (Ki67, 1:200 [ABCAM, Cambridge, Massachusetts]. Secondary antibodies (Invitrogen) were used at a 1:200 dilution for 1 hour at room temperature.

References

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100:** 3983-8.

Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB *et al* (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444:** 756-60.

Battula VL, Shi Y, Evans KW, Wang RY, Spaeth EL, Jacamo RO *et al* (2012). Ganglioside GD2 identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest*.

Baumann M, Krause M, Hill R (2008). Exploring the role of cancer stem cells in radioresistance. *Nat Rev Cancer* **8:** 545-54.

Binnerts ME, Tomasevic N, Bright JM, Leung J, Ahn VE, Kim KA *et al* (2009). The first propeller domain of LRP6 regulates sensitivity to DKK1. *Mol Biol Cell* **20:** 3552-60.

Bloom HJ, Richardson WW (1957). Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer* **11:** 359-77.

Brennan KR, Brown AM (2004). Wnt proteins in mammary development and cancer. *J Mammary Gland Biol Neoplasia* **9:** 119-31.

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T *et al* (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102:** 3738-43.

Chen B, Dodge ME, Tang W, Lu J, Ma Z, Fan CW *et al* (2009). Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat Chem Biol* **5:** 100-7.

Cho RW, Wang X, Diehn M, Shedden K, Chen GY, Sherlock G *et al* (2008). Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors. *Stem Cells* **26:** 364-71.

Clevers H The cancer stem cell: premises, promises and challenges. *Nat Med* **17:** 313-9.

Clevers H (2006). Wnt/β-catenin signaling in development and disease. *Cell* **127:** 469-80.

Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ (2005). Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res* **65:** 10946-51.

de Sousa EM, Vermeulen L, Richel D, Medema JP (2011). Targeting Wnt signaling in colon cancer stem cells. *Clin Cancer Res* **17:** 647-53.

Dodge ME, Lum L (2011). Drugging the cancer stem cell compartment: lessons learned from the hedgehog and Wnt signal transduction pathways. *Annu Rev Pharmacol Toxicol* **51:** 289-310.

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ *et al* (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* **17:** 1253-70.

Eguchi M, Nguyen C, Lee SC, Kahn M (2005). ICG-001, a novel small molecule regulator of TCF/β-catenin transcription. *Med Chem* **1:** 467-72.

Eramo A, Ricci-Vitiani L, Zeuner A, Pallini R, Lotti F, Sette G *et al* (2006). Chemotherapy resistance of glioblastoma stem cells. *Cell Death Differ* **13:** 1238-41.

Fang D, Nguyen TK, Leishear K, Finko R, Kulp AN, Hotz S *et al* (2005). A tumorigenic subpopulation with stem cell properties in melanomas. *Cancer Res* **65:** 9328-37.

Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M *et al* (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1:** 555-67.

Glinka A, Wu W, Delius H, Monaghan AP, Blumenstock C, Niehrs C (1998). Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391:** 357-62.

Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA *et al* (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138:** 645-59.

Guzman ML, Rossi RM, Karnischky L, Li X, Peterson DR, Howard DS *et al* (2005). The sesquiterpene lactone parthenolide induces apoptosis of human acute myelogenous leukemia stem and progenitor cells. *Blood* **105:** 4163-9.

Hallett RM, Dvorkin A, Gabardo CM, Hassell JA (2010). An algorithm to discover gene signatures with predictive potential. *J Exp Clin Cancer Res* **29:** 120.

Hirschhaeuser F, Menne H, Dittfeld C, West J, Mueller-Klieser W, Kunz-Schughart LA Multicellular tumor spheroids: an underestimated tool is catching up again. *J Biotechnol* **148:** 3-15.

Howe LR, Brown AM (2004). Wnt signaling and breast cancer. *Cancer Biol Ther* **3:** 36-41.

Khramtsov AI, Khramtsova GF, Tretiakova M, Huo D, Olopade OI, Goss KH Wnt/β-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am J Pathol* **176:** 2911-20.

Klarmann GJ, Decker A, Farrar WL (2008). Epigenetic gene silencing in the Wnt pathway in breast cancer. *Epigenetics* **3:** 59-63.

Kondratyev M, Kreso A, Hallett RM, Girgis-Gabardo A, Barcelon ME, Ilieva D *et al* (2011). Gamma-secretase inhibitors target tumor-initiating cells in a mouse model of ERBB2 breast cancer. *Oncogene* **31:** 93-103.

Korkaya H, Paulson A, Charafe-Jauffret E, Ginestier C, Brown M, Dutcher J *et al* (2009). Regulation of mammary stem/progenitor cells by PTEN/Akt/β-catenin signaling. *PLoS Biol* **7:** e1000121.

Kurpios NA, MacNeil L, Shepherd TG, Gludish DW, Giacomelli AO, Hassell JA (2009). The Pea3 Ets transcription factor regulates differentiation of multipotent progenitor cells during mammary gland development. *Dev Biol* **325:** 106-21.

Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J *et al* (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367:** 645-8.

Lepourcelet M, Chen YN, France DS, Wang H, Crews P, Petersen F *et al* (2004). Small-molecule antagonists of the oncogenic Tcf/β-catenin protein complex. *Cancer Cell* **5:** 91-102.

Lindvall C, Bu W, Williams BO, Li Y (2007). Wnt signaling, stem cells, and the cellular origin of breast cancer. *Stem Cell Rev* **3:** 157-68.

Lindvall C, Evans NC, Zylstra CR, Li Y, Alexander CM, Williams BO (2006). The Wnt signaling receptor Lrp5 is required for mammary ductal stem cell activity and Wnt1-induced tumorigenesis. *J Biol Chem* **281:** 35081-7.

Liu BY, McDermott SP, Khwaja SS, Alexander CM (2004). The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells. *Proc Natl Acad Sci U S A* **101:** 4158-63.

Liu CC, Prior J, Piwnica-Worms D, Bu G LRP6 overexpression defines a class of breast cancer subtype and is a target for therapy. *Proc Natl Acad Sci U S A* **107:** 5136-41.

Liu JC, Deng T, Lehal RS, Kim J, Zacksenhaus E (2007). Identification of tumorsphere- and tumor-initiating cells in HER2/Neu-induced mammary tumors. *Cancer Res* **67:** 8671-81.

Lopez-Knowles E, Zardawi SJ, McNeil CM, Millar EK, Crea P, Musgrove EA *et al* Cytoplasmic localization of β-catenin is a marker of poor outcome in breast cancer patients. *Cancer Epidemiol Biomarkers Prev* **19:** 301-9.

Ma S, Lee TK, Zheng BJ, Chan KW, Guan XY (2008). CD133+ HCC cancer stem cells confer chemoresistance by preferential expression of the Akt/PKB survival pathway. *Oncogene* **27:** 1749-58.

Meijer L, Skaltsounis AL, Magiatis P, Polychronopoulos P, Knockaert M, Leost M *et al* (2003). GSK-3-selective inhibitors derived from Tyrian purple indirubins. *Chem Biol* **10:** 1255-66.

Nusse R (2005). Wnt signaling in disease and in development. *Cell Res* **15:** 28-32.

O'Brien CA, Pollett A, Gallinger S, Dick JE (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445:** 106-10.

Pastrana E, Silva-Vargas V, Doetsch F (2011). Eyes wide open: a critical review of sphere-formation as an assay for stem cells. *Cell Stem Cell* **8:** 486-98.

Pullan SaCHS (1996). *Epithelial Cell Culture*. Cambridge University Press: Cambrudge

Reya T, Duncan AW, Ailles L, Domen J, Scherer DC, Willert K *et al* (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423:** 409-14.

Reynolds BA, Rietze RL (2005). Neural stem cells and neurospheres--re-evaluating the relationship. *Nat Methods* **2:** 333-6.

Reynolds BA, Weiss S (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **255:** 1707-10.

Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J *et al* (2003). Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63:** 5821-8.

Stingl J (2008). Detection and analysis of mammary gland stem cells. *J Pathol*.

Stingl J, Eaves CJ, Kuusk U, Emerman JT (1998). Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. *Differentiation* **63:** 201-13.

Suzuki H, Toyota M, Carraway H, Gabrielson E, Ohmura T, Fujikane T *et al* (2008). Frequent epigenetic inactivation of Wnt antagonist genes in breast cancer. *Br J Cancer* **98:** 1147-56.

Ugolini F, Charafe-Jauffret E, Bardou VJ, Geneix J, Adelaide J, Labat-Moleur F *et al* (2001). WNT pathway and mammary carcinogenesis: loss of expression of candidate tumor suppressor gene SFRP1 in most invasive carcinomas except of the medullary type. *Oncogene* **20:** 5810-7.

Vermeulen L, De Sousa EMF, van der Heijden M, Cameron K, de Jong JH, Borovski T *et al* (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat Cell Biol* **12:** 468-76.

Wong GT, Gavin BJ, McMahon AP (1994). Differential transformation of mammary epithelial cells by Wnt genes. *Mol Cell Biol* **14:** 6278-86.

Woodward WA, Chen MS, Behbod F, Alfaro MP, Buchholz TA, Rosen JM (2007). WNT/β-catenin mediates radiation resistance of mouse mammary progenitor cells. *Proc Natl Acad Sci U S A* **104:** 618-23.

Zhang M, Behbod F, Atkinson RL, Landis MD, Kittrell F, Edwards D *et al* (2008). Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. *Cancer Res* **68:** 4674-82.

Table and Figure legends

Table 1. Wnt/β-catenin pathway genes comprising the gene signature of differentially expressed genes between tumorspheres and mammospheres induced to differentiate.

Table 2. IC_{50} values (μM) for PKF118-310, PKF115-584, and CGP049090 in sphere forming assays.

Figure 1. Expression of Wnt pathway genes in tumorspheres (TMS), mammospheres (MMS) and mammospheres induced to differentiate (diffMMS). A) Heat maps of 3 independent TMS (A, n=3), MMS (B, n=3) and diffMMS (C, n=3) RNA preparations profiled on MOE430A Gene Chips. B-E) qRT-PCR of transcripts encoding components and target genes of the Wnt signaling pathway (* $p < 0.05$, t-test) for all genes shown. F and G) Survival, (*p=0.005, Log-rank test) and metastasis (*p=0.0068, Log-rank test), curves for human breast cancer patients classified based on their expression of TMS-related and TMS-unrelated (diffMMS-related) specific Wnt pathway components.

Figure 2. Agonists and antagonist of Wnt/β-catenin signaling regulate sphere and colony formation by primary tumor cells and primary mammary epithelial cells as well as by tumorsphere- and mammosphere-derived cells. A) Wnt3a and Dkk1 inhibit sphere formation by tumorsphere- and mammosphere-derived cells, compared to bovine serum albumin (BSA) and stem cell media (SCM) controls. B) BIO stimulates tumorsphere and mammosphere formation. C-D) Sphere formation in the presence of increasing concentrations of PKF118-310, PKF115-584, and CGP049090. E-F) Colony formation in the presence of increasing concentrations of PKF118-310, PKF115-584, and CGP049090.

Figure 3. PKF118-310 selectively targets breast TICs *in* vitro. A) PKF118-310 inhibits sphere formation by primary mammary epithelial cells. B) PKF118-310 inhibits sphere formation by primary tumor cells. C) PKF118-310 treated primary mammary epithelial cells form spheres with same efficiency compared to the DMSO vehicle. D) PKF118-310 treated primary tumor cells have little capacity to form spheres compared to the DMSO vehicle (* $p < 0.05$, t-test). E) Mass (g) of tumors formed from transplant of 1μM PKF118-310-treated primary tumor cells (* $p < 0.05$, t-test). F) Primary tumor cells treated with 2.5μM PKF118-310 are unable to initiate tumor growth after transplant into syngeneic recipient mice.

Figure 4. PKF118-310 treatment halts tumor growth. A) Tumor volumes of vehicle and PKF118-310 treated mice. B) Tumor volumes after completion of treatment (t=12 days) (*$p < 0.05$, t-test). C) PKF118-310 induces formation of duct-like structures (red arrows). D) Photographs of Ki67 stained tumor sections. E) PKF118-310 reduces the fraction of Ki67+ nuclei; quantification of Ki67-positive nuclei was assessed independently by two different individuals (*$p < 0.05$, t-test). F) Quantification of Wnt target gene expression in tumors isolated from treated and untreated mice using qRT-PCR (*$p < 0.05$, t-test).

Figure 5. Tumor cells from PKF118-310 treated tumor-bearing mice engraft and elicit the growth of secondary tumors less efficiently than their vehicle-treated counterparts. Viable tumor cells were harvested from PKF118-310- and DMSO-treated mice and 10,000 viable

tumor cells per mouse were transplanted into syngeneic recipients (n=20 treated/untreated)

(* $p < 0.05$, Log-rank test).

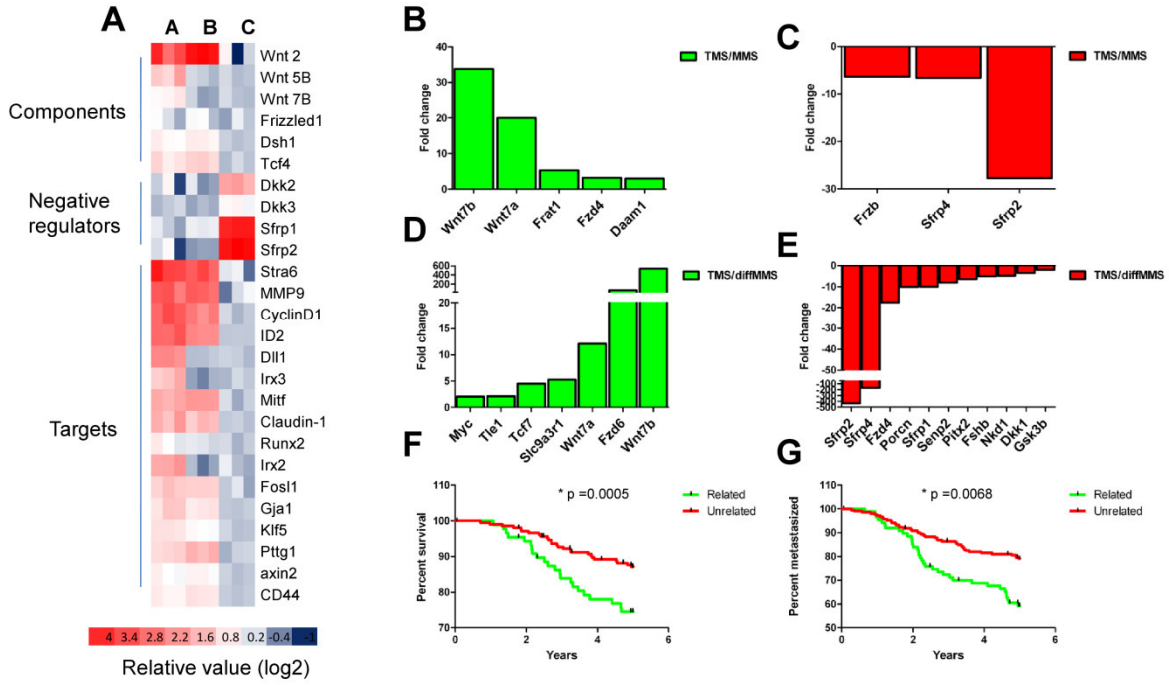
Table 1. Wnt/β-catenin pathway genes comprising the gene signature of differentially expressed genes between tumorspheres and mammospheres induced to differentiate

Expressed in tumorspheres	Expressed in mammospheres induced to differentiate
Wnt7A	**Sfrp1**
Myc	**Sfrp4**
Tle1	**Fzd4**
Tcf7	**Senp2**
SLC9A3R1	**Pitx2**
FZD6	**Fshb**
	Dkk1
	Gsk3β

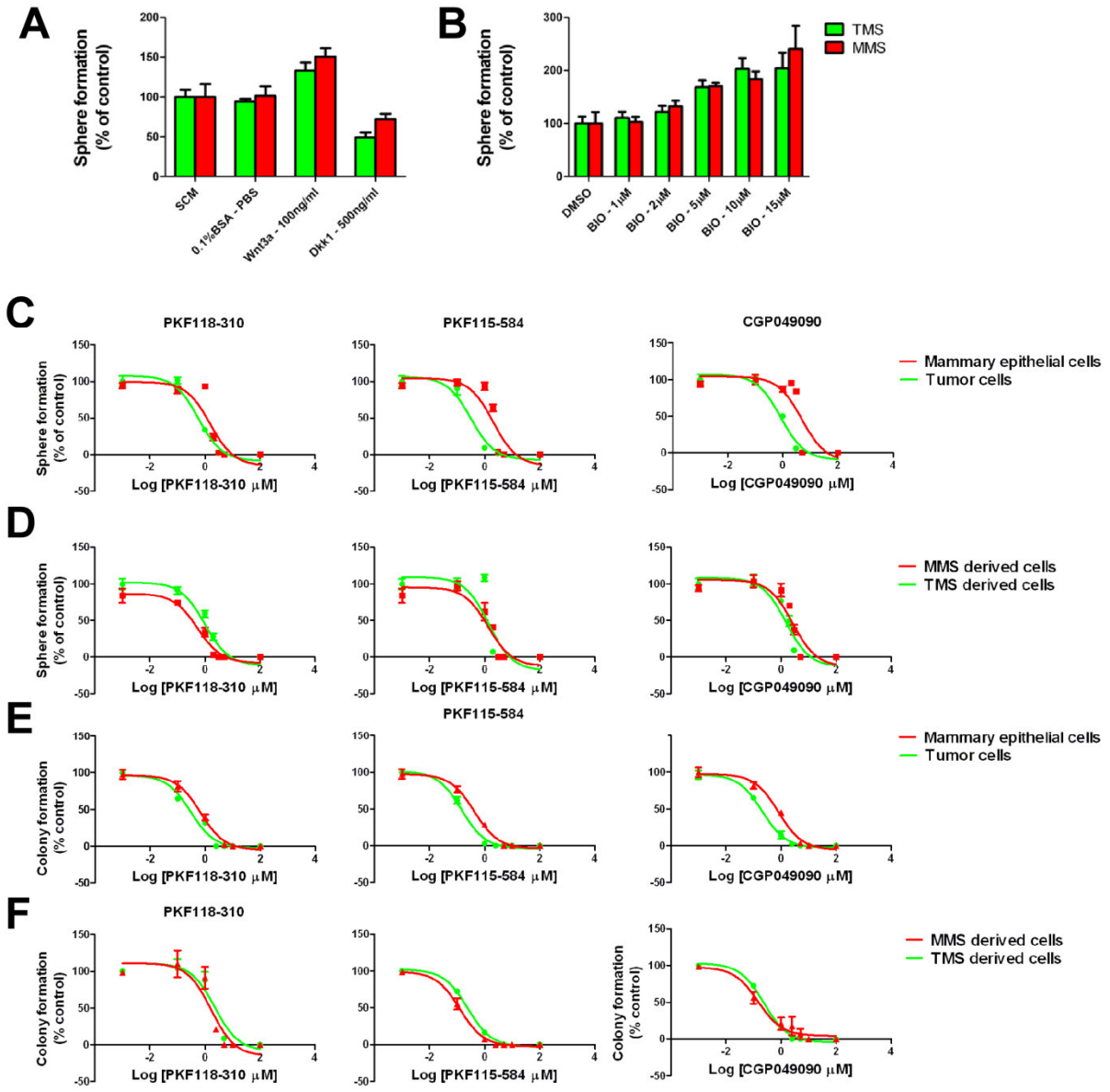
Table 2. IC_{50} values (μM) for PKF118-310, PKF115-584, and CGP049090 in sphere forming assays

Drug	Primary tumor Cells	Primary mammary epithelial cells	Tumorsphere-derived cells	Mammosphere-derived cells
PKF118-310	0.58	1.54	0.94	0.54
PKF115-584	0.31	2.05	1.34	1.39
CGP049090	0.84	4.89	1.52	2.64

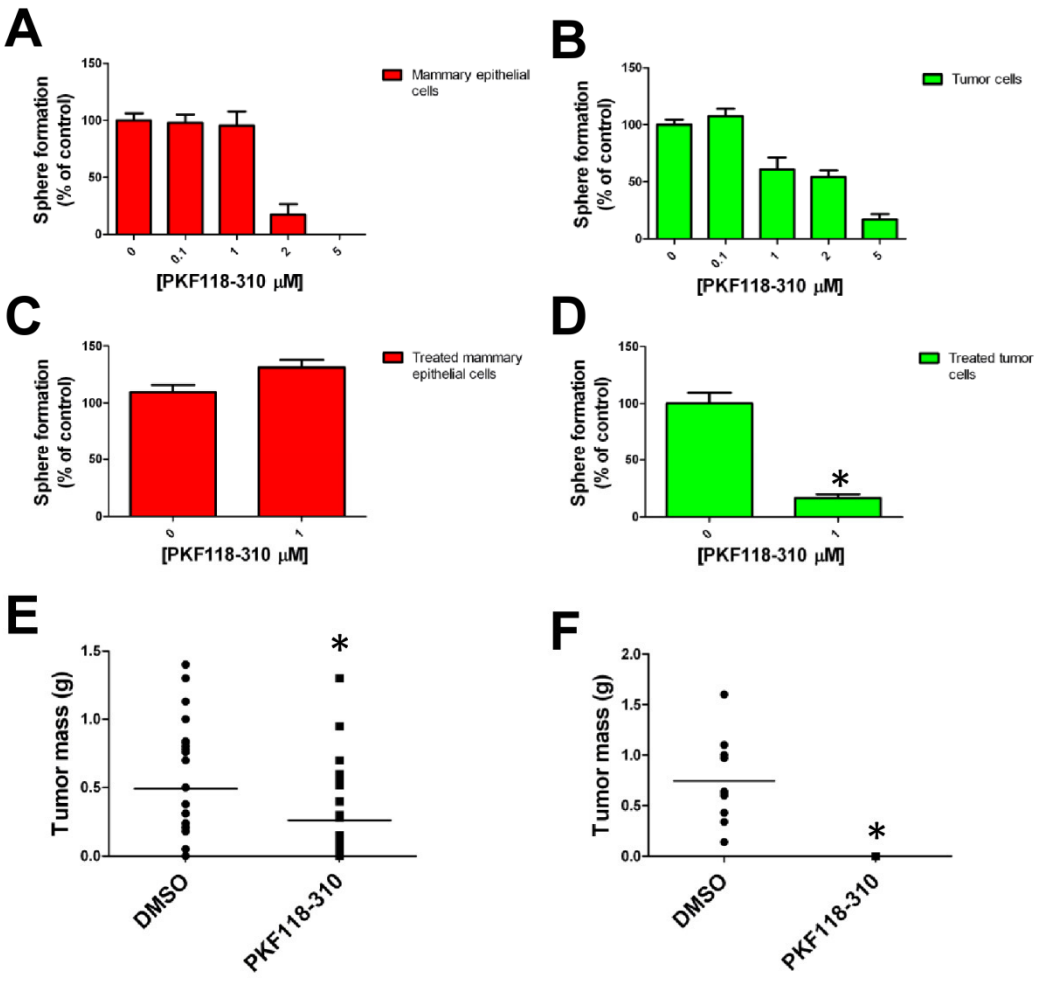
Hallett, R.M., et al – Figure 1

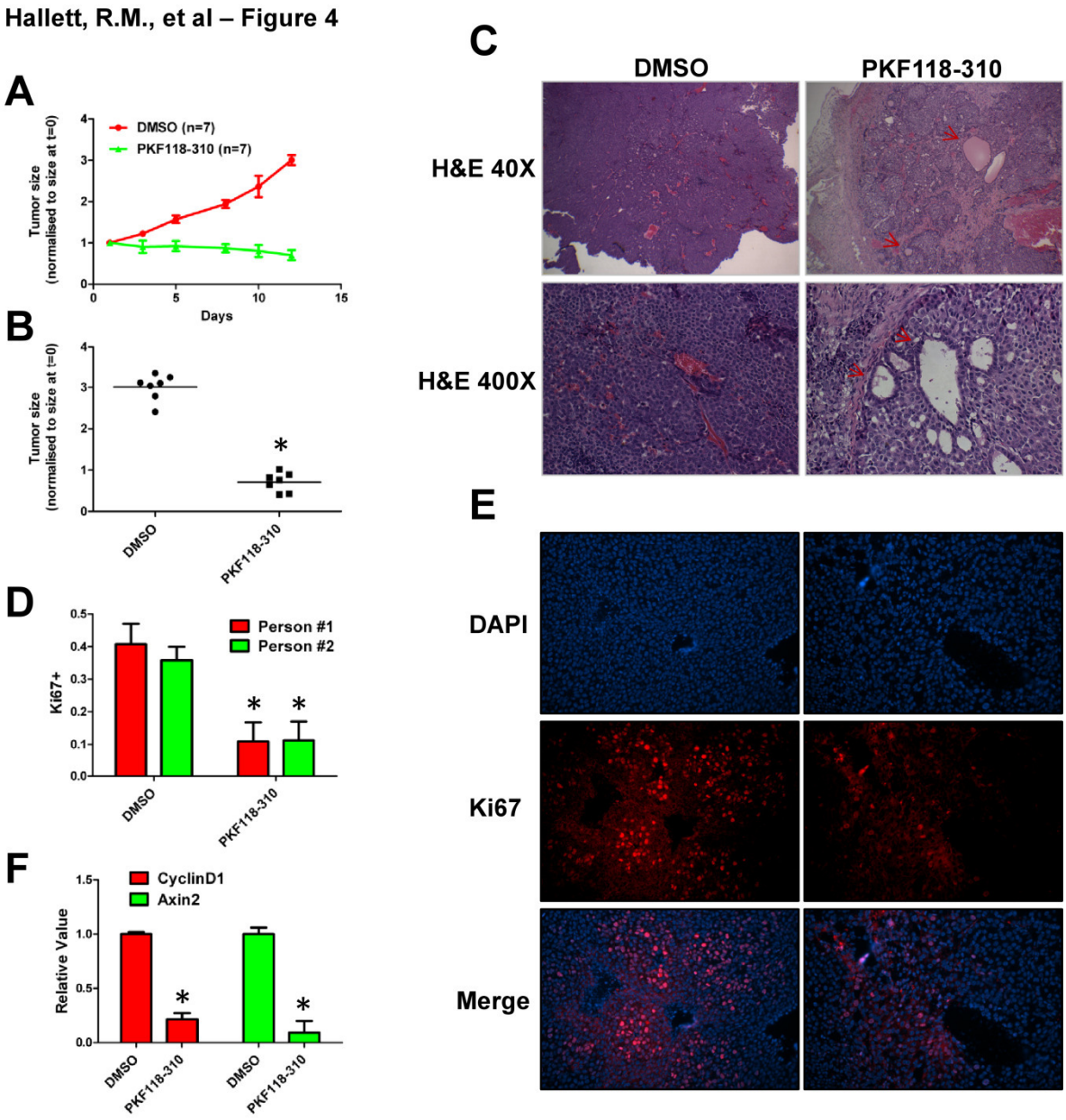


Hallett, R.M., et al – Figure 2

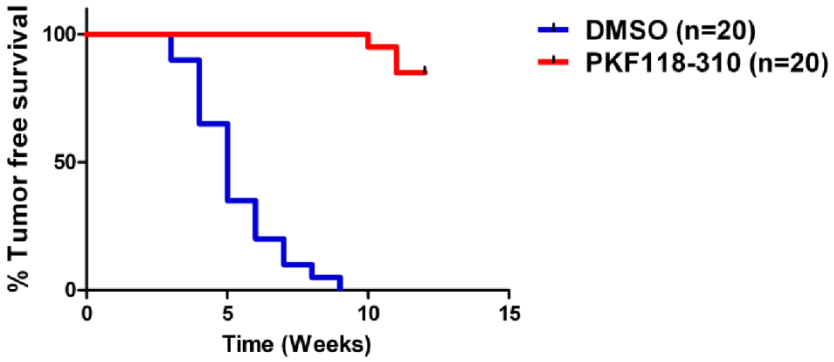


Hallett, R.M., et al – Figure 3




Hallett, R.M., et al – Figure 4
A
DMSO (n=7)
PKF118-310 (n=7)
Tumor size (normalised to size at t=0)
Days
B
Tumor size (normalised to size at t=0)
DMSO
PKF118-310
C
DMSO
PKF118-310
H&E 40X
H&E 400X
D
Person #1
Person #2
Ki67+
DMSO
PKF118-310
E
DAPI
Ki67
Merge
F
CyclinD1
Axin2
Relative Value
DMSO
PKF118-310
DMSO
PKF118-310

Hallett, R.M., et al – Figure 5



CHAPTER 7

Discussion

Multiple clinical trials suggest that commonly used treatments, such as tamoxifen and cytotoxic anti-mitotic therapeutics, increase long-term survival of breast cancer patients. However, it has become increasingly obvious that only a minor subset of patients benefit from the therapy under investigation. For example, the B-14 trial carried out by the National Surgical Adjuvant Breast and Bowel Project (NSABBP) found that adjuvant tamoxifen therapy increased 15-year recurrence-free survival from 65% to 78% (Fisher *et al.*, 2004). While this data strongly justifies the use of tamoxifen for patient treatment, it also highlights that the majority of patients do not benefit from adjuvant tamoxifen treatment. Notably, only 13% of patients derived tamoxifen-associated benefit, whereas of the remaining 87%, 65% did not need additional treatment post-surgery and 22% suffered a recurrence despite treatment post-surgery. Clinical trials examining the addition of cytotoxic chemotherapy to adjuvant tamoxifen produced similar outcomes. In the B-20 trial, which compared tamoxifen to tamoxifen plus cytotoxic chemotherapy, the rate of 12-year recurrence-free survival in tamoxifen-treated patients was 79% compared to 89% for patients treated with tamoxifen and chemotherapy (Fisher *et al.*, 2004). Again, this data justifies the use of chemotherapy in combination with tamoxifen for treating breast cancer, however they also highlight that the majority of patients do not benefit from adjuvant chemotherapy. Only 10% of patients derived benefit from adjuvant chemotherapy, whereas of the remaining 90%, 79% did not need require

chemotherapy and 11% suffered recurrence despite treatment. These data suggest that only approximately 1 in 10 patients derived benefit from adjuvant treatment post surgery and that among patients who received adjuvant therapy, the vast majority were over treated. Moreover, they also highlight that some patients were not afforded an effective therapy for treating their breast tumors. Taken together, these examples provide a compelling rationale for identifying biomarkers that could serve to identify patients who do not require adjuvant therapy as well as patients who will likely benefit from currently approved therapies. Furthermore, they also highlight the need to identify novel therapies for patients whose tumors are not durably eradicated by currently available therapies. Indeed, these challenges represent the overarching objectives of personalized or precision medicine for the treatment of breast cancer patients. The contents of this thesis describe experiments we performed to advance the goals of more personalized breast cancer medicine.

In chapter 2, we described the identification of a transcript-based gene signature that could identify BLBC patients with good and poor prognosis. Using publicly available data sets we identified 137 patients with BLBC for whom clinical follow-up data was also available, and identified a 14-gene signature whose expression was related to the clinical outcome of BLBC patients. Notably, in a validation cohort of node-negative chemotherapy-naïve patients, patients predicted to be high risk by the 14-gene signature experienced poor disease-free survival, and after 5 years approximately only 50% of patients remained disease free (Chapter 2, figure 3). In contrast, patients predicted to be low risk by the 14-gene signature, experienced excellent 5-year survival,

and no patients within this group suffered disease relapse within 5 years (Chapter 2, figure 3). Whereas these data strongly suggest that the 14-gene signature is related to the natural progression of BLBC, its capacity to identify good and poor outcome patients may in part be related to its capacity to predict patient response to chemotherapy. The relationship between our 14 gene signature and response to chemotherapy was not examined in our study. The capacity to guide prognosis in BLBC patients is particularly important because previous reports suggest that factors such as tumor size, tumor grade, extent of vascular invasion, and patient age, which represent the standard clinical variables used for patient prognosis, show little relationship to patient outcome in the context of BLBC, especially in lymph node-negative patients (Hudis and Gianni; Rakha *et al.*, 2007a). The only standard clinical variable that is consistently prognostic in BLBC appears to be nodal status (Hernandez-Aya *et al.*; Rakha *et al.*, 2007a). Based on these data, it seems likely that the biological programs that underpin the natural progression of BLBCs are different than those for other subtypes of breast cancer. In support of this, we found that with the exception of the HER2/ERBB2+ subgroup, the 14-gene signature was not related to prognosis in most other subtypes of breast cancer (Chapter 2, figure 5). Moreover, we found that multiple gene signatures, which were generally identified in patient cohorts comprising a majority of ER+ breast tumors, were not prognostic in the context of BLBC (Chapter 2, figure 4). Other studies suggest that the prognostic power of most signatures is derived from their capacity to measure proliferation and estrogen-receptor signaling (Desmedt *et al.*, 2008; Wirapati *et al.*, 2008). Whereas these biological programs are strongly linked to patient outcome in ER+ cancers, they fail to discriminate

between good and poor prognosis patients with ER- tumors, such as BLBC. Moving forwards, we are working to validate these findings using patient samples obtained from Hamilton Health Sciences, and eventually aim to complete a clinical trial using a refined version of our BLBC prognosis signature to guide the use of chemotherapy for patients with BLBC. Positive findings from such a trial would provide a compelling rational for the clinical implementation of our BLBC prognosis signature, such that BLBC patients predicted to have an excellent prognosis could avoid adjuvant treatment with chemotherapy whereas aggressive therapy could be given to those predicted to suffer disease relapse.

Whereas chapter 2 dealt with the identification of genes that could stratify BLBC into good and poor prognosis disease groups, in chapter 3 we sought to define distinct disease subgroups of ER+ breast cancers. Based on molecular subtype classification, the majority of ER+ breast tumor classify into the luminal A and luminal B subtypes (Perou *et al.*, 2000; Sorlie *et al.*, 2001; Sorlie *et al.*, 2003); however, more recent reports suggest that this molecular stratification fails to account for the complete heterogeneity observed among various ER+ tumors (Gatza *et al.*, 2010). Improved classification of breast tumors into therapeutically relevant subgroups provides a means to treat individuals with subgroup specific therapies comprising agents to which the tumor is most likely to respond. We hypothesized that investigation and characterization of the estrogen independent biology that underpins ER+ tumors might provide a means to stratify ER+ tumors into different distinct disease subgroups. We observed that the expression of estrogen independent genes could be used to assign ER+ into 1 of 8 different subgroups.

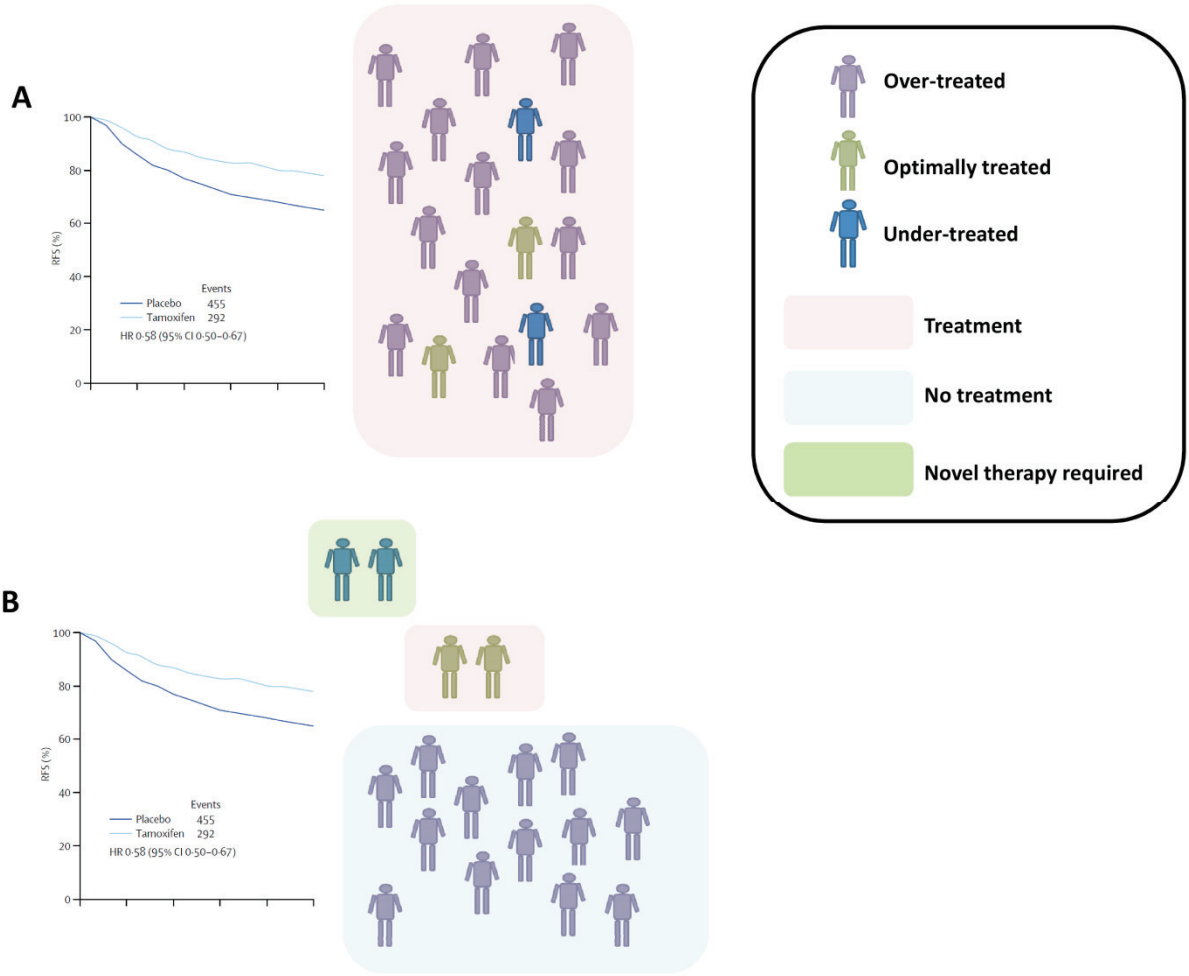
Importantly, these 8 subgroups were re-identified in a relatively large validation cohort, and the patients comprising these subgroups experienced different survival characteristics, suggesting that the tumor subgroups represent distinct disease entities (Chapter 3, figures 1&2). Whereas the knowledge of the molecular subtypes of breast cancer has existed for more than a decade, this knowledge has yet to translate into subtype-specific therapies for breast cancer patients. In the context of ER+ breast cancer, it is known that luminal B tumors are more susceptible to chemotherapy, whereas luminal A tumors rarely respond to chemotherapy (Hatzis *et al.*, 2011; Rouzier *et al.*, 2005a). To learn whether the 8 subgroups of ER+ tumors we identified could inform therapy selection we compared the response of cell lines representative of these subgroups to experimental and clinical compounds. Subgroup responses to commonly used chemotherapeutic agents, such as anthracycline and taxane drugs, were concordant between cell lines and patient tumors (Chapter 3, figure 3). We also identified several additional compounds which showed subgroup specific activity (Chapter 3, figure 4). We believe such data could be rapidly implemented into clinical practice, were upon clinical presentation of ER+ breast cancer, a biopsy is used for gene expression profiling analysis to make a subgroup assignment. Low risk patients would only receive endocrine therapy, whereas high risk patients would receive endocrine therapy as well as additional therapy informed by the subgroup of the tumor. However, significant additional work would need to be carried out to achieve these objectives, such as the development of a robust clinical multi-gene assay to optimally assign subgroup membership. Moreover,

the predictive power of subgroup guided decision making would need to be validated in prospective clinical trials.

Many studies have explored using the expression of targets as a means of predicting response to a given target's drug counterpart. This concept is supported by the observation that breast cancer patients whose tumors express the ER are more likely to respond to tamoxifen, a drug that targets the ER, than those whose tumors do not express the ER. However, well validated biomarkers that guide the use of conventional cytotoxic chemotherapy agents do not exist. Notably, *TOP2A or β-tubulin,* the targets of anthracycline and taxane drugs, have only been modestly linked to anthracycline and taxane response in previous studies, respectively (Desmedt *et al.*, 2011; Kavallaris *et al.*, 1997; Mozzetti *et al.*, 2005; Slamon and Press, 2009; Tanner *et al.*, 2006; Tommasi *et al.*, 2007). We hypothesized that measurement of target as well as target-associated transcripts in a tumor sample might provide a more comprehensive measure of molecular target activity, and thus the tumor's likely response to therapy. To this end we identified *TOP2A* and *β-tubulin* transcript indices, and tested their capacity to identify patients who were likely or unlikely to respond to chemotherapy regimens that comprised anthracycline and taxane drugs (Chapter 4, figure 2&3). Excitingly, these indices were statistically associated with chemotherapy responses in a multivariate model that included several additional clinical variables, such as ER status and grade, suggesting that these indices could be used to inform treatment selection for breast cancer patients, and imply that these indices could be used to guide the use of anthracycline and taxane drugs in breast cancer patients.

Taken together, the experiments discussed above were designed to identify new biomarkers that could be used to maximize the effectiveness of currently available breast therapeutics. Based on clinical trial data (Fisher *et al.*, 2004), the implementation of such biomarkers has the potential to avoid overtreatment of as many as ~60% of breast cancer patients, and properly target treatment to ~30% of the remaining ~40% of patients (Discussion figure below). Unfortunately, no effective therapy exists to provide durable cure to the remaining ~10% of patients. Hence new therapies are needed that are effective in this population of available-treatment-refractory tumors.

As described in the introduction, many studies have identified subpopulations of cells within tumors, termed cancer stem cells or tumor-initiating cells (TIC), which are the suspected culprits responsible for driving tumor relapse (Al-Hajj *et al.*, 2003; Clevers; Collins *et al.*, 2005; Fang *et al.*, 2005; Lapidot *et al.*, 1994; Lobo *et al.*, 2007; O'Brien *et al.*, 2007; Prince *et al.*, 2007; Reya *et al.*, 2001; Singh *et al.*, 2003). Based on these data, it seems logical that therapies that target breast TIC would likely benefit patients who do not receive durable cure from current therapies. We took multiple approaches to identify such therapies, including the use of breast cancer prone mice that develop tumors enriched in breast TICs, as well as human breast tumors and breast tumor cell lines.



Discussion figure. A) When breast cancer patients are treated identically, the majority are over-treated, and a minority are either optimally treated or under-treated. **B)** Biomarkers, such as those described in this thesis, could serve to identify low risk patients (basal signature, ER+ subgroups) as well as patients that would benefit from currently approved therapies (ER+ subgroups, *TOP2A* and *β-tubulin* indices). Moreover, patients identified as high risk, but unlikely to respond to current therapeutic options could be enrolled in clinical trials testing experimental treatments, such as those that target TIC (nAChR/Wnt pathway inhibitors).

The first approach to identify therapies that target breast TICs was to use the response of bulk tumor cells to chemotherapy as a surrogate marker of breast TICs. Since breast TICs are resistant to chemotherapy (Gong *et al.*, 2009; Gupta *et al.*, 2009; Lacerda *et al.*, 2010; Lobo *et al.*, 2007), it seems likely that breast tumors and breast tumor cell lines that are enriched in TICs would be less susceptible to chemotherapy. Using this approach, we identified genes whose expression was related to chemotherapy response in a large cohort of breast cancer patients as well as breast tumor cell lines. Based on the identified transcriptional models of chemotherapy resistance, we identified the small molecule adiphenine, which increased the expression of genes associated with chemotherapy sensitivity whereas it decreased the expression of chemotherapy resistant genes. Importantly, we found that adiphenine as well as its analogs could induce sensitivity to chemotherapy in breast tumor cell lines that were otherwise resistant to these drugs, suggesting that adiphenine and analogs thereof targeted TICs in bulk human breast tumor cell cultures, thus accounting for their capacity to overcome chemotherapy resistance in human breast tumors (Chapter 5, figure 3). Moreover, adiphenine and analogs inhibited sphere formation by several human breast tumor cell lines, and reduced the frequency of TICs as assayed by transplant into immuno-compromised mice (Chapter 5, figures 4&5). Whereas the identity of cells that form spheres is not known with certainty, mammosphere and tumorsphere forming cells co-fractionate with mammary epithelial stem/progenitor cells, as well as breast TIC, respectively (Liu *et al.*, 2007a; Pastrana *et al.*, 2011). Agents affecting the frequency of sphere-forming cells

correspondingly alter the frequency of mammary epithelial stem/progenitor cells, as well as breast TIC, as measured by cell transplantation experiments (Kondratyev *et al.*, 2011), suggesting that sphere formation is a reasonable surrogate assay for measuring TICs. Taken together, these data are consistent with the hypothesis that adiphenine and analogs target breast TICs. Excitingly, since adiphenine is an old drug (Gaensler and McGowan, 1950), previously used as an anti-spasmolytic, the development of adiphenine as an anticancer drug could likely be rapidly expedited. That said, additional work remains to be completed, such as testing the *in vivo* efficacy of adiphenine or analogs of adiphenine in human breast tumor xenografts, to validate our observations made *in vitro*.

Whereas we used chemotherapy resistance in bulk tumor cell preparations as a surrogate marker of breast TICs to identify adiphenine, we also completed experiments with relatively pure populations of breast TICs harvested from mouse mammary tumors. We compared the gene expression profiles of tumorspheres derived from MMTV-*Neu* tumors with strain matched (FVB/N) mammospheres to identify signaling pathways which might be hyperactive in breast TICs relative to their normal stem/progenitor cell counterparts (Chapter 6, figure 1). With this approach, we observed that components and target genes of the Wnt/β-catenin singaling pathway were expressed at higher levels in tumorspheres than in mammospheres, suggesting that inhibition of this pathway might selectively target breast TICs. Notably, several studies support the inhibition of Wnt/β-catenin as a promising approach for breast cancer treatment (Khramtsov *et al.*; Korkaya *et al.*, 2009; Liu *et al.*, 2010; Suzuki *et al.*, 2008). For our experiments we used the chemical inhibitor PKF118-310 to inhibit Wnt/β-catenin pathway activity in a variety of

assays designed to measure its effects on breast TICs. PKF118-310 is a reported inhibitor of Wnt/β-pathway signaling that interrupts the protein-protein interaction between β-catenin and TCF4 (Lepourcelet *et al.*, 2004). The best evidence that PKF118-310 targets breast TICs is that when we transplanted tumor cells isolated from tumor bearing mice treated with PKF118-310, these cells either failed or elicited tumor growth after long-latency compared to tumor cells isolated from vehicle treated mice (Chapter 6, figure 5), suggesting that PKF118-310 dramatically reduced the frequency of functional breast TICs *in vivo*. Indeed, we believe our studies with PKF118-310 provide a strong rationale for testing PKF118-310 in human xenografts models of breast cancer as well as for the development of Wnt/β-catenin pathway inhibitors for breast cancer therapy.

The central tenet of personalize breast cancer medicine is to treat patients based on the characteristics of their tumors. In this thesis we describe the identification of several biomarker gene signatures, which following suitable clinical development and optimization could be used to spare breast cancer patients from over-treatment, as well as treat those patients who require treatment with the optimal approved therapy. Finally, for the subset of patients who require additional treatment, but whose tumors are unlikely to respond to approved therapies, we describe the identification of experimental agents that target breast TICs and could serve as a starting point to develop more effective therapeutics for these patients

REFERENCES

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100:** 3983-8.

Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K *et al* (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* **22:** 2284-93.

Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB *et al* (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444:** 756-60.

Battula VL, Shi Y, Evans KW, Wang RY, Spaeth EL, Jacamo RO *et al* (2012). Ganglioside GD2 identifies breast cancer stem cells and promotes tumorigenesis. *J Clin Invest*.

Baumann M, Krause M, Hill R (2008). Exploring the role of cancer stem cells in radioresistance. *Nat Rev Cancer* **8:** 545-54.

Bear HD, Anderson S, Smith RE, Geyer CE, Jr., Mamounas EP, Fisher B *et al* (2006). Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer:National Surgical Adjuvant Breast and Bowel Project Protocol B-27. *J Clin Oncol* **24:** 2019-27.

Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D *et al* (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439:** 353-7.

Bogaerts J, Cardoso F, Buyse M, Braga S, Loi S, Harrison JA *et al* (2006). Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol* **3:** 540-51.

Borst P, Wessels L (2010). Do predictive signatures really predict response to cancer chemotherapy? *Cell Cycle* **9:** 4836-40.

Bruce WR, Van Der Gaag H (1963). A Quantitative Assay for the Number of Murine Lymphoma Cells Capable of Proliferation in Vivo. *Nature* **199:** 79-80.

Bueno-de-Mesquita JM, van Harten WH, Retel VP, van't Veer LJ, van Dam FS, Karsenberg K *et al* (2007). Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol* **8:** 1079-87.

Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM *et al* (2006). Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* **98:** 1183-92.

Carraway KL, Price-Schiavi SA, Komatsu M, Jepson S, Perez A, Carraway CA (2001). Muc4/sialomucin complex in the mammary gland and breast cancer. *J Mammary Gland Biol Neoplasia* **6:** 323-37.

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T *et al* (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102:** 3738-43.

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL *et al* (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10:** 529-41.

Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP *et al* (2007). High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol* **8:** R215.

Cho RW, Wang X, Diehn M, Shedden K, Chen GY, Sherlock G *et al* (2008). Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors. *Stem Cells* **26:** 364-71.

Clevers H The cancer stem cell: premises, promises and challenges. *Nat Med* **17:** 313-9.

Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ (2005). Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res* **65:** 10946-51.

Colombo PE, Milanezi F, Weigelt B, Reis-Filho JS (2011). Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. *Breast Cancer Res* **13:** 212.

Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM *et al* (2005). A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res* **65:** 4059-66.

Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J *et al* (2011). Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol* **29:** 1578-86.

Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G *et al* (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* **14:** 5158-65.

Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P *et al* (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med* **17:** 1086-93.

Eramo A, Ricci-Vitiani L, Zeuner A, Pallini R, Lotti F, Sette G *et al* (2006). Chemotherapy resistance of glioblastoma stem cells. *Cell Death Differ* **13:** 1238-41.

Fang D, Nguyen TK, Leishear K, Finko R, Kulp AN, Hotz S *et al* (2005). A tumorigenic subpopulation with stem cell properties in melanomas. *Cancer Res* **65:** 9328-37.

Fisher B, Costantino J, Redmond C, Poisson R, Bowman D, Couture J *et al* (1989). A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med* **320:** 479-84.

Fisher B, Dignam J, Wolmark N, DeCillis A, Emir B, Wickerham DL *et al* (1997). Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst* **89:** 1673-82.

Fisher B, Jeong JH, Bryant J, Anderson S, Dignam J, Fisher ER *et al* (2004). Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet* **364:** 858-68.

Gaensler EA, McGowan JM (1950). Activity of some newer spasmolytic agents on the biliary tract of man; a comparative study of isopropyl-norepinephrine, adiphenine hydrochloride, methylamino-iso-octene and khellin. *J Pharmacol Exp Ther* **99:** 304-11.

Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD *et al* (2010). A pathway-based classification of human breast cancer. *Proc Natl Acad Sci U S A* **107:** 6994-9.

Ghods AJ, Irvin D, Liu G, Yuan X, Abdulkadir IR, Tunici P *et al* (2007). Spheres isolated from 9L gliosarcoma rat cell line possess chemoresistant and aggressive cancer stem-like cells. *Stem Cells* **25:** 1645-53.

Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M *et al* (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1:** 555-67.

Gong C, Yao H, Liu Q, Chen J, Shi J, Su F *et al* (2009). Markers of tumor-initiating cells predict chemoresistance in breast cancer. *PLoS One* **5:** e15630.

Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C *et al* (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146:** 633-44.

Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA *et al* (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138:** 645-59.

Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M *et al* (2006). A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* **8:** R25.

Hallett RM, Dvorkin-Gheva A, Anita B, Hassell JA (2012). A gene signature for predicting outcome in patients with basal-like breast cancer. *Scientific Reports* **2**.

Hallett RM, Dvorkin A, Gabardo CM, Hassell JA (2010). An algorithm to discover gene signatures with predictive potential. *J Exp Clin Cancer Res* **29:** 120.

Hallett RM, Hassell JA (2011). E2F1 and KIAA0191 expression predicts breast cancer patient survival. *BMC Res Notes* **4:** 95.

Hamburger AW, Salmon SE (1977). Primary bioassay of human tumor stem cells. *Science* **197:** 461-3.

Haraguchi N, Utsunomiya T, Inoue H, Tanaka F, Mimori K, Barnard GF *et al* (2006). Characterization of a side population of cancer cells from human gastrointestinal system. *Stem Cells* **24:** 506-13.

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S *et al* (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* **25:** 5287-312.

Hassane DC, Sen S, Minhajuddin M, Rossi RM, Corbett CA, Balys M *et al* Chemical genomic screening reveals synergism between parthenolide and inhibitors of the PI-3 kinase and mTOR pathways. *Blood* **116:** 5983-90.

Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A *et al* (2011). A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* **305:** 1873-81.

Heppner GH (1984). Tumor heterogeneity. *Cancer Res* **44:** 2259-65.

Herbst RS, Bajorin DF, Bleiberg H, Blum D, Hao D, Johnson BE *et al* (2006). Clinical Cancer Advances 2005: major research advances in cancer treatment, prevention, and screening--a report from the American Society of Clinical Oncology. *J Clin Oncol* **24:** 190-205.

Hernandez-Aya LF, Chavez-Macgregor M, Lei X, Meric-Bernstam F, Buchholz TA, Hsu L *et al* Nodal status and clinical outcomes in a large cohort of patients with triple-negative breast cancer. *J Clin Oncol* **29:** 2628-34.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z *et al* (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8:** R76.

Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA *et al* (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24:** 4236-44.

Holland DG, Burleigh A, Git A, Goldgraben MA, Perez-Mancera PA, Chin SF *et al* (2011). ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol Med* **3:** 167-80.

Holstege H, Horlings HM, Velds A, Langerod A, Borresen-Dale AL, van de Vijver MJ *et al* BRCA1-mutated and basal-like breast cancers have similar aCGH profiles and a high incidence of protein truncating TP53 mutations. *BMC Cancer* **10:** 654.

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF *et al* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7:** 96.

Hudis CA, Gianni L Triple-negative breast cancer: an unmet medical need. *Oncologist* **16 Suppl 1:** 1-11.

Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J *et al* (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* **66:** 10292-301.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011). Global cancer statistics. *CA Cancer J Clin* **61:** 69-90.

Kang MK, Kang SK (2007). Tumorigenesis of chemotherapeutic drug-resistant cancer stem-like cells in brain glioma. *Stem Cells Dev* **16:** 837-47.

Kavallaris M, Kuo DY, Burkhart CA, Regl DL, Norris MD, Haber M *et al* (1997). Taxol-resistant epithelial ovarian tumors are associated with altered expression of specific beta-tubulin isotypes. *J Clin Invest* **100:** 1282-93.

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* **176:** 2911-20.

Kondratyev M, Kreso A, Hallett RM, Girgis-Gabardo A, Barcelon ME, Ilieva D *et al* (2011). Gamma-secretase inhibitors target tumor-initiating cells in a mouse model of ERBB2 breast cancer. *Oncogene*.

Korkaya H, Paulson A, Charafe-Jauffret E, Ginestier C, Brown M, Dutcher J *et al* (2009). Regulation of mammary stem/progenitor cells by PTEN/Akt/beta-catenin signaling. *PLoS Biol* **7:** e1000121.

Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H *et al* (2007). Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* **9:** R65.

Kwei KA, Kung Y, Salari K, Holcomb IN, Pollack JR Genomic instability in breast cancer: pathogenesis and clinical implications. *Mol Oncol* **4:** 255-66.

Lacerda L, Pusztai L, Woodward WA (2010). The role of tumor initiating cells in drug resistance of breast cancer: Implications for future therapeutic approaches. *Drug Resist Updat* **13:** 99-108.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ *et al* (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313:** 1929-35.

Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J *et al* (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367:** 645-8.

Lee-Hoeflich ST, Crocker L, Yao E, Pham T, Munroe X, Hoeflich KP *et al* (2008). A central role for HER3 in HER2-amplified breast cancer: implications for targeted therapy. *Cancer Res* **68:** 5878-87.

Lee JK, Coutant C, Kim YC, Qi Y, Theodorescu D, Symmans WF *et al* Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer. *Clin Cancer Res* **16:** 711-8.

Lepourcelet M, Chen YN, France DS, Wang H, Crews P, Petersen F *et al* (2004). Small-molecule antagonists of the oncogenic Tcf/beta-catenin protein complex. *Cancer Cell* **5:** 91-102.

Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA *et al* (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* **26:** 1275-81.

Liedtke C, Wang J, Tordai A, Symmans WF, Hortobagyi GN, Kiesel L *et al* (2010). Clinical evaluation of chemotherapy response predictors developed from breast cancer cell lines. *Breast Cancer Res Treat* **121:** 301-9.

Liu CC, Prior J, Piwnica-Worms D, Bu G (2010). LRP6 overexpression defines a class of breast cancer subtype and is a target for therapy. *Proc Natl Acad Sci U S A* **107:** 5136-41.

Liu JC, Deng T, Lehal RS, Kim J, Zacksenhaus E (2007a). Identification of tumorsphere- and tumor-initiating cells in HER2/Neu-induced mammary tumors. *Cancer Res* **67:** 8671-81.

Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T *et al* (2007b). The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* **356:** 217-26.

Lobo NA, Shimono Y, Qian D, Clarke MF (2007). The Biology of Cancer Stem Cells. *Annu Rev Cell Dev Biol*.

Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C *et al* (2007). Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* **25:** 1239-46.

Lu Y, Zi X, Pollak M (2004). Molecular mechanisms underlying IGF-I-induced attenuation of the growth-inhibitory activity of trastuzumab (Herceptin) on SKBR3 breast cancer cells. *Int J Cancer* **108:** 334-41.

Lu Y, Zi X, Zhao Y, Mascarenhas D, Pollak M (2001). Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin). *J Natl Cancer Inst* **93:** 1852-7.

Ma S, Lee TK, Zheng BJ, Chan KW, Guan XY (2008). CD133+ HCC cancer stem cells confer chemoresistance by preferential expression of the Akt/PKB survival pathway. *Oncogene* **27:** 1749-58.

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY *et al* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133:** 704-15.

Mozzetti S, Ferlini C, Concolino P, Filippetti F, Raspaglio G, Prislei S *et al* (2005). Class III beta-tubulin overexpression is a prominent mechanism of paclitaxel resistance in ovarian cancer patients. *Clin Cancer Res* **11:** 298-305.

Nagata Y, Lan KH, Zhou X, Tan M, Esteva FJ, Sahin AA *et al* (2004). PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* **6:** 117-27.

Nalwoga H, Arnes JB, Wabinga H, Akslen LA (2008). Expression of EGFR and c-kit is associated with the basal-like phenotype in breast carcinomas of African women. *APMIS* **116:** 515-25.

Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z *et al* (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* **10:** 5367-74.

O'Brien CA, Pollett A, Gallinger S, Dick JE (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445:** 106-10.

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M *et al* (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351:** 2817-26.

Paik S, Tang G, Shak S, Kim C, Baker J, Kim W *et al* (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* **24:** 3726-34.

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T *et al* (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27:** 1160-7.

Pastrana E, Silva-Vargas V, Doetsch F (2011). Eyes wide open: a critical review of sphere-formation as an assay for stem cells. *Cell Stem Cell* **8:** 486-98.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA *et al* (2000). Molecular portraits of human breast tumours. *Nature* **406:** 747-52.

Pierce GB, Johnson LD (1971). Differentiation and cancer. *In Vitro* **7:** 140-5.

Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R *et al* (2006). Genomic signatures to guide the use of chemotherapeutics. *Nat Med* **12:** 1294-300.

Potti A, Nevins JR (2008). Utilization of genomic signatures to direct use of primary chemotherapy. *Curr Opin Genet Dev* **18:** 62-7.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI *et al* (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* **12:** R68.

Prince ME, Sivanandan R, Kaczorowski A, Wolf GT, Kaplan MJ, Dalerba P *et al* (2007). Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc Natl Acad Sci U S A* **104:** 973-8.

Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO (2007a). Prognostic markers in triple-negative breast cancer. *Cancer* **109:** 25-32.

Rakha EA, El-Sayed ME, Green AR, Paish EC, Powe DG, Gee J *et al* (2007b). Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. *J Clin Oncol* **25:** 4772-8.

Reis-Filho JS, Westbury C, Pierga JY (2006). The impact of expression profiling on prognostic and predictive testing in breast cancer. *J Clin Pathol* **59:** 225-31.

Reya T, Morrison SJ, Clarke MF, Weissman IL (2001). Stem cells, cancer, and cancer stem cells. *Nature* **414:** 105-11.

Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E *et al* (2009). T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res* **11:** R15.

Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K *et al* (2005a). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* **11:** 5678-85.

Rouzier R, Pusztai L, Delaloge S, Gonzalez-Angulo AM, Andre F, Hess KR *et al* (2005b). Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. *J Clin Oncol* **23:** 8331-9.

Rutgers E, Piccart-Gebhart MJ, Bogaerts J, Delaloge S, Veer LV, Rubio IT *et al* (2011). The EORTC 10041/BIG 03-04 MINDACT trial is feasible: results of the pilot phase. *Eur J Cancer* **47:** 2742-9.

Sabatier R, Finetti P, Mamessier E, Raynaud S, Cervera N, Lambaudie E *et al* Kinome expression profiling and prognosis of basal breast cancers. *Mol Cancer* **10:** 86.

Sachlos E, Risueno RM, Laronde S, Shapovalova Z, Lee JH, Russell J *et al* (2012). Identification of Drugs Including a Dopamine Receptor Antagonist that Selectively Target Cancer Stem Cells. *Cell*.

Scaltriti M, Rojo F, Ocana A, Anido J, Guzman M, Cortes J *et al* (2007). Expression of p95HER2, a truncated form of the HER2 receptor, and response to anti-HER2 therapies in breast cancer. *J Natl Cancer Inst* **99:** 628-38.

Shoemaker RH (2006). The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* **6:** 813-23.

Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J *et al* (2003). Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63:** 5821-8.

Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E *et al* (2011). ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol Med* **3:** 153-66.

Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE *et al* (1989). Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244:** 707-12.

Slamon DJ, Press MF (2009). Alterations in the TOP2A and HER2 genes: association with adjuvant anthracycline sensitivity in human breast cancers. *J Natl Cancer Inst* **101:** 615-8.

Sorlie T (2004). Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Eur J Cancer* **40:** 2667-75.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98:** 10869-74.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100:** 8418-23.

Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A *et al* (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* **100:** 10393-8.

Sotiriou C, Pusztai L (2009). Gene-expression signatures in breast cancer. *N Engl J Med* **360:** 790-800.

Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J *et al* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* **98:** 262-72.

Spillane JB, Henderson MA (2007). Cancer stem cells: a review. *ANZ J Surg* **77:** 464-8.

Staaf J, Jonsson G, Ringner M, Vallon-Christersson J, Grabau D, Arason A *et al* (2010). High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. *Breast Cancer Res* **12:** R25.

Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J *et al* (2001). Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* **98:** 10787-92.

Suzuki H, Toyota M, Carraway H, Gabrielson E, Ohmura T, Fujikane T *et al* (2008). Frequent epigenetic inactivation of Wnt antagonist genes in breast cancer. *Br J Cancer* **98:** 1147-56.

Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V *et al* (2007). Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol* **25:** 4414-22.

Tanner M, Isola J, Wiklund T, Erikstein B, Kellokumpu-Lehtinen P, Malmstrom P *et al* (2006). Topoisomerase IIalpha gene amplification predicts favorable treatment response to tailored and dose-escalated anthracycline-based adjuvant chemotherapy in HER-2/neu-amplified breast cancer: Scandinavian Breast Group Trial 9401. *J Clin Oncol* **24:** 2428-36.

Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C (2007). An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol* **8:** R157.

Tommasi S, Mangia A, Lacalamita R, Bellizzi A, Fedele V, Chiriatti A *et al* (2007). Cytoskeleton and paclitaxel sensitivity in breast cancer: the role of beta-tubulins. *Int J Cancer* **120:** 2078-85.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M *et al* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530-6.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L *et al* (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* **20:** 719-26.

Vuaroqueaux V, Urban P, Labuhn M, Delorenzi M, Wirapati P, Benz CC *et al* (2007). Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome. *Breast Cancer Res* **9:** R33.

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F *et al* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365:** 671-9.

Weigelt B, Baehner FL, Reis-Filho JS (2010). The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol* **220:** 263-80.

Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B *et al* (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res* **10:** R65.

Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B (2001). Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. *J Natl Cancer Inst Monogr***:** 96-102.

Wulf GG, Wang RY, Kuehnle I, Weidner D, Marini F, Brenner MK *et al* (2001). A leukemic stem cell with intrinsic drug efflux capacity in acute myeloid leukemia. *Blood* **98:** 1166-73.

Yau C, Esserman L, Moore DH, Waldman F, Sninsky J, Benz CC (2010). A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *Breast Cancer Res* **12:** R85.

Zhang M, Behbod F, Atkinson RL, Landis MD, Kittrell F, Edwards D *et al* (2008). Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. *Cancer Res* **68:** 4674-82.