Germline and Somatic Variation in Genes Involved in Metabolism of Chemotherapy Drugs and Breast Cancer Outcomes

By

Michelle L Baglia

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in
Epidemiology
May, 2016
Nashville, Tennessee

Approved:

Xiao-Ou Shu, M.D., Ph.D., M.P.H.

Jirong Long, Ph.D.

Fei Ye, Ph.D., M.S.P.H.

Sarah Nechuta, Ph.D., M.P.H.

To Mark, Carys, Adrian, and Isabelle,
For their never-ending love and support,

To my Mom and Dad,
For always believing in me

ACKNOWLEDGEMENTS

I want to thank Dr. Xiao-Ou Shu, my mentor and dissertation committee chair, for sharing her infinite wisdom with me. I am grateful that she saw potential in me and allowed me to complete my PhD under her mentorship. I have learned a great deal from her over the past four years. She has always kept both my career and life aspirations in mind as she has guided me throughout this process and on to the next phase of my career. I also want to thank each of my committee members, Dr. Jirong Long, Dr. Fei Ye, and Dr. Sarah Nechuta. I am fortunate to have been able to complete a module with each one of them in order to flesh out methods for different aspects of my dissertation. I am grateful that each of them has taken the time to build my knowledge base and help me to grow into a stronger researcher. All of my committee members have provided me with timely and comprehensive feedback on my work which has been integral to completing my dissertation. Additionally, I would like to thank my professors and other researchers in my department; for their advice on research projects, career decisions, and life in general. I am grateful for all I have learned from each of them. I would also like to thank the research staff and participants of the Shanghai Breast Cancer Survival Study for making this work possible. My doctoral studies were also supported financially through the Vanderbilt CTSA award from the National Center for Advancing Translational Sciences.

This work could not have been completed without the love and support of my friends and family. Thank you all for being there when I needed to talk things out, de-stress, or vent. My parents have always believed in me and constantly remind me that I can do anything. Their never-ending, and sometimes outrageous, confidence in me has been a driving force in all aspects of my life.

It would not be possible to put into words the gratitude that I have for my husband, Mark, for his love and support during this process. He has always supported me even when my path was not clear. He has spent countless hours talking through my research with me and helping me to keep my life in perspective. I could not have done this without him. I am also thankful to my children, Carys, Adrian, and Isabelle, for always giving me a good reason to step away from my research for a few hours and for reminding me that I am even stronger than I ever thought possible. I hope that they remember this time in our lives and learn that they too can do anything if they are willing to work for it.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Epidemiology of Triple-Negative Breast Cancer

Triple-negative breast cancer (TNBC) is an aggressive breast cancer subtype which is characterized by minimal or no expression of estrogen receptors (ER) and progesterone receptors (PR), and the absence of overexpression of human epidermal growth factor 2 (HER2). In the United States, TNBC makes up 15-20% of breast cancers.[1-3] These tumors tend to occur more in women who are young and/or African-American.[4] Additionally, they are typically diagnosed at later stages than other breast cancer subtypes and are associated with poorer prognosis.[5] Due to its aggressive nature, rates of recurrence and mortality are higher in TNBC compared with other breast cancer subtypes and prognosis by breast cancer subtype varies by time since diagnosis.[6,7]

Current Treatment for TNBC

The primary treatment for TNBC consists of combination chemotherapy. Despite that studies have found that TNBC tumors are more sensitive to chemotherapy than hormone receptor positive cancers,[3,4,8-12] a pathologic complete response (pCR) only occurs in ~30% of TNBC cases.[13] One study showed that there was no difference in 3-yr survival between TNBC and non-TNBC patients when a pCR was achieved, but survival was significantly lower for TNBC patients as compared to non-TNBC patients who did not achieve pCR (68% vs 88%, p=0.001).[10] Residual disease is common following treatment and accounts for worse outcomes and an increased risk of metastatic recurrence;[9,13] therefore, differences in survival between TNBC and

non-TNBC patients may be due to the presence of residual disease, as well as the aggressiveness of this tumor subtype.[14]

Although chemotherapy is the standard treatment, currently TNBC does not have a standard treatment plan for chemotherapy type/combination and/or dosage. TNBC tumors treated with neoadjuvant anthracycline- or anthracycline plus taxane-based chemotherapy regimens have been shown to respond well and have higher rates of pCR compared to ER+ cancers; however, in those who do not achieve pCR, relapse rates are high.[9,11,15] Other studies have also shown that TNBC tumors respond better to antimetabolite-based regimens, such as cyclophosphamide, methotrexate, and 5-fluorouracil (CMF) or doxorubicin and cyclophosphamide (AC), compared to other tumor types.[15,16] Additionally, other studies have shown that, compared to conventional chemotherapy agents, platinum compounds may be more effective for TNBC tumors.[15] Radiotherapy is also used in the treatment of TNBC; although chemotherapy is generally considered to be the more effective component of treatment.[11,15]

Due to a lack of consensus in the medical literature as to which provides the best response in TNBC tumors;[3,4,17,18] the chemotherapy regimen that a TNBC patient receives often depends upon practices of the patient's physician and treatment hospital.[4] The molecular mechanisms which underlie response to chemotherapy and achieving pCR are not understood. Research aimed at identifying predictors for treatment response and survival is critically important for delivering effective treatments to TNBC patients with minimal toxicity.

Drug Metabolism of Chemotherapeutic Agents

The efficacy and cytotoxic effects of chemotherapeutic agents are largely determined by the metabolism and transport of the agents in tumor cells and non-tumor tissue.[19] Drug sensitivity and resistance are controlled by drug metabolism and pharmacogenetic variability in metabolism may account for differences in treatment efficacy.[20] The plasma concentration of a drug over time determines the pharmacological effect with too little exposure being associated with an ineffective treatment while too much may lead to adverse events, including toxicity.[21] Individual differences in drug response and cancer survival may be predictable with a better understanding of the genetic variability in drug metabolizing genes.[21]

Metabolism of Cyclophosphamide

Cyclophosphamide belongs to the class of chemotherapy drugs called alkylating agents and is used in the treatment of various cancers.[22] It undergoes both activating and inactivating metabolic reactions in the body.[23] Cyclophosphamide is activated to 4-hydroxycyclophosphomide through catalyzation by cytochrome P450 (CYP) isozymes in the liver; the primary metabolizing enzymes are CYP2B6, CYP2C9, and CYP3A4 while CYP3A5, CYP2A6, CYP2C8, and CYP2C19 metabolize cyclophosphamide to a lesser extent.[24-26] 4-hydroxycyclophosphamide rapidly interconverts to aldophosphamide; both molecules are thought to passively leave liver cells, circulate, and passively enter other cells, including tumor cells.[24,27] Therefore, individual variation in CYP metabolism likely plays a role in the amount of the circulating active metabolite of cyclophosphamide which is available to enter tumor cells. Aldophosphamide yields phosphoramide mustard, likely the clinically important DNA cross-linking agent, through a spontaneous elimination reaction, which is associated with bladder

toxicity.[24] Phosphoramide mustard does not readily enter cells in its anionic form so the intracellular formation of this metabolite from aldophosphamide is important.[24] Aldehyde dehydrogenase 1A1 (ALDH1A1) is primarily responsible for the detoxification of aldophosphamide, with aldehyde dehydrogenase 3A1 (ALDH3A1) involved to a lesser extent.[23,28] Both 4-hydroxycyclophosphamide and phosphoramide mustard are detoxified by glutathione S-transferases (GSTs); GSTA1 and GSTP1 are the main isoforms involved, while GSTT1 and GSTM1 are involved to a lesser extent.[23]

Germline Genetic Variability in Cyclophosphamide Metabolism

Genetic variability in genes involved in the metabolism of cyclophosphamide has been shown to affect breast cancer outcomes following treatment, although the evidence is not entirely consistent.[20,29-31] We have reviewed all of the available evidence regarding the SNPs of interest and metabolism of cyclophosphamide and all cancer types due to the paucity of information in breast cancer alone.

CYP2B6

A cohort study of 230 breast cancer patients (97% Caucasian) treated with cyclophosphamide and doxorubicin observed shorter progression-free survival and overall survival (OS) in patients with certain germline genotypes of two variants, rs12721655 and rs3745274, in CYP2B6.[16] However, a clinical trial in 882 breast cancer patients (83% Caucasian) found no association between the rs3745274 polymorphism and disease-free survival (DFS) in breast cancer patients treated with cyclophosphamide, 5-fluorouracil, and methotrexate or doxorubicin versus those who did not receive chemotherapy treatment.[32] The rs3211371 polymorphism in the CYP2B6 gene has also been shown to be associated with

leucopenia/neutropenia resulting in dose delay and dose reduction.[16] A study in 107 leukemia patients treated with cyclophosphamide found certain CYP2B6 variants (rs8192709, rs3745274, and rs2279343) were associated with risk of adverse side effects.[33]

CYP3A4

One SNP in the CYP3A4 gene has been previously identified as playing a role in response to cyclophosphamide treatment.[34] In that study, 127 premenopausal breast cancer cases (89% Caucasian) treated with a cyclophosphamide-based chemotherapy regimen found an increased risk of ovarian failure among young women (<45 years) with the variant alleles in the rs2740574 SNP compared to wild type alleles.[34] Additionally, another study showed that slower metabolism of cyclophosphamide is associated with shorter survival times in 85 chemotherapy-naïve breast cancer patients (87% Caucasian) and that variant genotypes of the rs2740574 polymorphism of the CYP3A4 gene were associated with slower drug activation and decreased survival.[35]

CYP2C19

In the CYP2C19 gene, one SNP (rs4244285) was found to be associated with overall survival in 230 breast cancer patients (97% Caucasian) receiving cyclophosphamide treatment in combination with doxorubicin, with reduced survival observed for individuals that were homozygous for the variant allele.[16] This indicates that those carrying the variant allele are slower metabolizers than those homozygous for the wild type allele.

ALDH1A1/ALDH3A1

Five SNPs in the ALDH1A1 gene have been suggested to play a role in the inactivation of aldophosphamide, a metabolite of cyclophosphamide, in cancer patients. In 882 breast cancer patients (83% Caucasian) enrolled in a clinical trial who were treated with cyclophosphamide and doxorubicin, three SNPs in the ALDH1A1 gene (rs3764435, rs8187996, and rs63319) were associated with drug toxicity before correcting for multiple comparisons ($p<0.05$).[36] However, none remained significant when a corrected *p*-value was used. The two SNP haplotype (rs3764435-rs168351) was associated with increased drug toxicity and this remained significant after correction for multiple comparisons.[36] Another SNP in the ALDH1A1 gene, rs6151031, was associated with drug toxicity in a small study of 113 Caucasian cancer patients, mostly breast cancer, treated with cyclophosphamide, thiotepa, and carboplatin.[37] This study also identified a SNP in the ALDH3A1 gene (rs2228100) as associated with likelihood of developing cystitis.[37]

GSTP1/GSTA1/GSTM1/GSTT1

In a study of 87 Yakut ovarian cancer patients treated with cyclophosphamide and cisplatin, the SNP, rs3957357, in GSTA1 was shown to be associated with anemia, but not with other adverse side effects, including neutropenia.[38] These authors previously reported a significant association between this polymorphism and overall survival in a population of 104 Russian ovarian cancer patients.[39] However, the authors found no association between this variant and overall or progression-free survival in 87 Yakut ovarian cancer patients.[38] Furthermore, when the 104 Russian patients were reanalyzed with additional follow-up time, the association between this polymorphism and overall survival was no longer significant.[38] This

study also found no association between GSTP1 SNP, rs1138272, and drug toxicity, progression-free, or overall survival in these ovarian cancer patients.[38]

Another functional SNP, rs1695, in the GSTP1 gene has been shown to be associated with increased drug response[40,41] and increased severity of toxicity[40] in populations of breast cancer patients in China (n=122)[40] and Brazil (n=40)[41] treated with cyclophosphamide and epirubicin. This SNP was shown to be associated with progression-free survival in a population of 104 Russian ovarian cancer patients treated with cyclophosphamide and cisplatin;[39] however, this association was not observed when that cohort was reanalyzed with additional follow-up time or in similar patients of Yakut ethnicity (n=87).[38]

No studies were identified which assessed the association between single nucleotide polymorphisms in the GSTM1 or GSTT1 genes and cancer outcomes among those who underwent chemotherapy with cyclophosphamide. Some studies have shown that deletion of the GSTT1 gene,[42] GSTM1 gene,[35,43] or both[44,45] was associated with a significantly better breast cancer prognosis. However, several other studies have shown no association with either the deletion of GSTT1 or GSTM1 and breast cancer outcomes[20,41,46,47] and one study showed an association between deletion of the GSTM1 gene and poorer breast cancer survival.[48]

Metabolism of 5-Fluorouracil

5-fluorouracil belongs to the class of chemotherapy drugs called antimetabolites and is one of the most frequently used chemotherapy drugs for solid tumors, including breast cancer.[49,50] The efficacy of 5-fluorouracil depends on the activation of the drug to the active metabolite in the body and the subsequent deactivation.[51] 5-fluorouracil is activated

intracellularly into several active metabolites by thymidine phosphorylase (TYMP) and uridine monophosphate synthetase (UMPS) (also known as orotate phosphoribosyltransferase (OPRT)).[52] The active metabolites inhibit the conversion from deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP).[52,53] It does this by targeting the enzyme, thymidylate synthase (TS), encoded by the TYMS gene, which is responsible for catalyzing the methylation of dUMP to dTMP.[54] This step is critical for DNA replication of the cell. Thus, the active metabolites of 5-fluorouracil block dTMP production in cancer cells which prevents DNA synthesis and replication. This step occurs when the active metabolite forms a stable complex with TS, along with folate as a co-factor.[55,56] Due to the role of folate as a co-factor in this step, methylenetetrahydrofolate reductase (MTHFR), which metabolizes folate and forms the co-factor required for inhibition of TS, has been suggested as potential factor in fluorouracil metabolism.[57,58] Reduced MTHFR activity, resulting from mutations in the MTHFR gene, may increase the rate of activity of TS due to the resulting increased folate metabolites.[57] Mutations in the TYMS gene may result in reduced expression of TS which may be associated with greater response to 5-fluorouracil.

Dihydropyrimidine dehydrogenase (DPD), an enzyme encoded by the DPYD gene, is the rate-limiting step in 5-fluorouracil metabolism and is responsible for greater than 80% of the degradation and inactivation of 5-fluorouracil.[52,59] This gene has been widely implicated as a strong predictor of fluorouracil response.[49,57,60,61] Studies have shown that reduced DPYD activity may lead to increased efficacy and/or increased risk of toxicity due to slower clearance of the drug and accumulation of dUMP.[57] Degradation of 5-fluorouracil by DPYD is highest in the liver but occurs in all tissues, including tumor tissue.[62,63]

Germline Genetic Variability in 5-Fluorouracil Metabolism

Genetic variability in genes involved in the metabolism of 5-fluorouracil has been shown to affect breast cancer outcomes following treatment, although the evidence is not entirely consistent.[29-31]

DPYD

The primary gene of interest in 5-fluorouracil metabolism is DPYD due to its role in degrading and inactivating greater than 80% of 5-fluorouracil. However, the association between DPYD polymorphisms and clinical outcomes is not fully understood due to the rarity and heterogeneity of functional variants in different populations.[64,65] The majority of research on genetic variability, 5-fluorouracil response, and cancer outcomes has been done in colorectal cancer (CRC) patients due to the impact of this drug in treatment of CRC; however, it is commonly used in breast cancer patients as well.[52] Additionally, the majority of research on this gene has been conducted in primarily Caucasian populations. There are three SNPs in the DPYD gene—rs67376798, rs3918290, and rs55886062—for which there is a high level of evidence that variant alleles are associated with decreased DPYD activity and increased risk of toxicity.[24] The results from these studies are summarized below.

In studies of CRC patients receiving fluorouracil in combination with other chemotherapy agents, the rs67376798 SNP was associated with altered DPYD activity,[66] decreased clearance of fluorouracil,[67] and increased risk or severity of toxicity,[67-72] though some studies have shown no association with toxicity.[73] A meta-analysis of 7 studies evaluating the association between this SNP and drug toxicity in cancer (primarily CRC) patients treated with 5-fluorouracil (either alone or in combination with other chemotherapy drugs) found an

increased risk of severe toxicity in patients carrying at least one variant allele.[74] The majority of the patients included in these studies were Caucasian.

The rs3918290 SNP has been shown to be associated with altered DPD activity and decreased 5-fluorouracil clearance,[67,75] increased risk of adverse side effects,[57,70,76,77] and increased risk of severity of toxicity[48,67-69,71,72,78-81] among cancer patients receiving combination chemotherapy including 5-fluorouracil. However, some studies which investigated this polymorphism did not observe an association with drug toxicity.[73,82,83] In a meta-analysis of 13 studies, which included various types of cancer patients who were treated with 5-fluorouracil (in combination or alone), an association was found between the variant allele in this SNP and risk of drug toxicity.[74] One study found an increased risk of death in cancer patients treated with 5-fluorouracil who carried a variant allele[84] and another study, in one patient heterozygous for the variant, also showed an association with death.[85] A study in various types of cancer patients treated with fluorouracil showed no association between this variant and risk of death; however, only two heterozygotes were included in the study and these patients did not experience fatal toxicity.[86] In all of the studies, very few, if any, homozygotes for the rs3918290 variant allele were observed; these studies found increased risk associated with carrying only one variant allele.

The rs55886062 SNP has been shown to be associated with DPYD deficiency[87] and risk of severity of toxicity[72] in cancer patients (various types, including breast cancer) who received fluorouracil. Decreased DPD activity was also observed in heterozygotes from a population of healthy European-Americans.[88]

Numerous other alleles in the DPYD gene have been shown to be associated with response to fluorouracil though the evidence for these polymorphisms is not as strong. In populations of patients with various cancer types (primarily CRC) treated with fluorouracil, it has been suggested that the following polymorphisms are associated with drug toxicity, rs115232898,[89] rs17376848,[48] rs1801158,[90,91] rs1801160,[76] rs2297595,[69,91,92] rs45589337,[91] rs56038477,[73] or adverse side effects, rs1801159,[93] rs1801265[93] (although this polymorphism was found to have no association with toxicity in other studies[67,72,91]), and rs75017182[94] (an association was found in CRC patients but another study which included various cancer types found no association with toxicity[95]). Additionally, the following polymorphisms have been shown to be associated with DPD activity in healthy populations, rs115232898,[88] rs115632870,[88] rs1801160,[88] and rs72728438,[88] and in cancer patients, rs1801158[87]. Other polymorphisms in the DPYD gene have not been found to be associated with response or toxicity to 5-fluorouracil—rs17116806[96] and rs4970722[96]—although the studies on these polymorphisms are very limited.

TYMS

It has been suggested that polymorphisms in the TYMS gene may be associated with response to 5-fluorouracil. The rs2847153 was found to be associated with survival in 211 pancreatic cancer patients (93% Caucasian) treated with fluorouracil.[97] The rs34489327 6-base pair insertion/deletion polymorphism was found to be associated with risk of disease progression in a population of 146 Caucasian CRC patients treated with fluorouracil, leucovorin, and irinotecan.[98] The rs34743033 tandem repeat polymorphism was associated with gene expression of TYMS[99,100] and toxicity[48,101] and response[99,102-106] to fluorouracil in cancer patients (various

types); the number of repeats was found to be associated with gene expression and outcomes; although the evidence was not consistent.[107-110]

MTHFR

MTHFR genetic variants, namely rs1801131 and rs1801133, lead to decreased enzyme activity in vitro although studies have shown inconsistent results in terms of effects on tumor response to chemotherapy treatment, disease progression, and survival.[51,53,64,111] The rs1801131 polymorphism has been shown to be associated with response to chemotherapy treatment in CRC patients treated with fluorouracil, in combination with other chemotherapy drugs.[104,112] Similarly, the rs1801133 polymorphism has been shown to be associated with drug toxicity[108,110] and neutropenia[113] in cancer patients (primarily CRC) treated with fluorouracil, although not all studies have observed an association.[48,95] The majority of this research on the association between 5-fluorouracil and MTHFR has been conducted in CRC; however, a previous study from our research group showed that the TT genotype of the MTHFR 677C>T polymorphism (rs1801133) was associated with shorter survival in Chinese breast cancer patients with advanced stage disease treated with chemotherapy.[114] Another Spanish study of MTHFR in 93 breast cancer patients showed no association with disease-free survival for this polymorphism.[115] It should be noted that the majority of patients included in the previous studies were of European descent.

TYMP

The rs11479 polymorphism has been shown to be associated with response to 5-fluorouracil and capecitabine. In a study of 253 CRC patients treated with 5-fluorouracil or

capecitabine, those who carried at least one A allele were at a significantly increased risk of toxicity as compared to those that were homozygous for the G allele.[94]

UMPS

Three polymorphisms in the UMPS gene have been evaluated for their associations with cancer outcomes in a population of 89 CRC patients with liver metastases treated with 5-fluorouracil or capecitabine.[116] Those who carried one or more A alleles in the rs2291078 polymorphism had a significantly worse response to chemotherapy as compared to those that were homozygous for the T allele. In the rs3772809 polymorphism, those patients with one or more A alleles had a significantly improved response to chemotherapy as compared to those who were homozygous for the G allele. In the rs3772810 polymorphism, those patients with one or more A alleles had a significantly improved response to chemotherapy as compared to those who were homozygous for the G allele.

Somatic Variation in Chemotherapy Metabolizing Genes

It has been suggested that somatic variation in genes involved in chemotherapy metabolism within the tumor may account for differences in response to chemotherapy and cancer prognosis. Few studies have evaluated tumor-level expression of genes involved in chemotherapy metabolism and response to chemotherapy or prognosis among breast cancer patients. While several reviews on germline genetic variability in chemotherapy metabolizing genes and breast cancer outcomes have been conducted,[23,29,30] no systematic reviews or meta-analyses summarizing previous research on tumor-level gene or protein expression of chemotherapy metabolizing genes in breast cancer patients were found.

Tumor Expression of Chemotherapy Metabolizing Genes and Breast Cancer Outcomes: A Systematic Review

Due to the known toxic effects and variation in tumor response to chemotherapy and cancer prognosis, there is a continued need for prognostic and predictive biomarkers to guide clinical decision making in chemotherapy drug selection. It has been suggested that somatic variations in chemotherapy metabolizing genes within the tumor may account for differences in response to chemotherapy and cancer prognosis. We performed a systematic search of epidemiologic studies investigating associations between expression of chemotherapy metabolizing genes and breast cancer outcomes. We included any survival outcomes, including overall, disease-free, and progression-free survival, and response outcomes, including pCR and response rate. We focused on those classes of genes known to metabolize the majority of chemotherapy drugs. Full methods for the systematic review can be found in Appendix 1. A total of 15 studies were included in the review (Table 1).

Table 1: Summary of Study Characteristics and Results

Author	Study Location	Years of diagnosis	Number of participants	Length of follow-up (median(range))	Age at diagnosis (median (range) or mean (SD))	Exposure(s)	Outcomes	Chemotherapy	Unit of Analysis	Univariate results	Multivariable results
Cytochrome P450 Genes											
Chang et al[117] (2008)	USA, UK	2000-2004	72	n/a	49 (not reported)	CYP1B1 gene expression	CR	Docetaxel	gene expression (per unit increase)	Higher expression associated with increased likelihood of CR: OR=1.70, 95% CI: 1.02, 2.95, *p*=0.0421 per 1-unit increment in expression	None reported for CYP1B1
Kolacinska et al[118] (2012)	Poland		42	n/a	55.6 (32-80)	CYP2D6 gene expression	pCR	Doxorubicin and docetaxel (n=29), doxorubicin and cyclophosphamide (n=13)	gene expression	ANOVA: CYP2D6 expression higher among those who achieve pCR compared to partial and non-responders (*p*=0.0063)	None reported
Gianni et al[119] (2005)	Italy, USA	1998-2002	Italy: n=89; USA: n=82	n/a	49.9 (29-65)	CYP3A4 and ALDH1A1 gene expression	pCR	Neoadjuvant doxorubicin and paclitaxel, adjuvant CMF	gene expression (per unit increase)	Italy: per unit increase CYP3A4: OR=1.82, 95% CI: 0.94-3.52, *p*=0.0462; ALDH1A1: OR=2.12, 95% CI: 0.93-4.81, *p*=0.0415; USA: no specific data for genes of interest	None reported for individual genes of interest
Aldehyde Dehydrogenase Genes											
Khoury et al[120] (2012)	USA	1995-2007	513	Not reported	Not reported; ≤50 (n=190) >50 (n=323)	ALDH1A1 via Tissue Microarray	DFS and OS	None (n=29); Neoadjuvant cyclophosphamide (n=36); Adjuvant cyclophosphamide (n=377) in combination with methotrexate and 5-FU or adriamycin with or without taxane	ALDH1A1-positive (≥10% tumor cell staining) vs ALDH1A1-negative (<10% staining)	KM survival curves, All patients: ALDH1A1 significantly associated with worse OS (*p*=0.04) but not DFS	Patients treated in neoadjuvant setting with cyclophosphamide without trastuzumab: OS (ALDH1A1+ vs -) HR=11.56 (2.13, 62.86) *p*=0.005, DFS HR=3.05 (0.85, 10.93) *p*=0.09; Patients treated in the neoadjuvant setting without trastuzumab: OS HR=7.08 (1.61, 31.13) *p*=0.01, DFS HR=2.57 (0.78, 8.52)

Table 1 continued

Author	Study Location	Years of diagnosis	Number of participants	Length of follow-up (median(range))	Age at diagnosis (median (range) or mean (SD))	Exposure(s)	Outcomes	Chemotherapy	Unit of Analysis	Univariate results	Multivariable results
Liu et al[121] (2015)	China, USA	China: 2002-2006 USA: 2001-2011	China: 463; USA: 133	Not reported	China: 51.6 (26.1-74.3) USA: 54 (28-75)	ALDH1A1 gene expression	DFS and OS	China: mostly CMF or CEF, no chemo (n=28); USA: unknown	Gene expression	KM survival curves: China: High ALDH1A1 expression associated with better DFS (p=0.01) and better OS (p=0.048)	Per unit increase ALDH1A1 expression: China: DFS: HR=0.87, 95% CI: 0.80, 0.95; OS: HR=0.85, 95% CI: 0.78-0.93); USA: OS: HR=0.88, 95% CI: 0.79, 0.93
Nogami et al[122] (2014)	Japan	1998-2006	40	46 (6-143) months	53 (28-78)	ALDH1A1 protein expression	DFS	Adjuvant chemotherapy (either anthracycline or anthracycline + taxane (n=10) or CMF (n=3))	ALDH1 positive (>5% tumor cells stained) vs negative	KM survival curves/Univariate Cox models: No association with disease-free survival for ALDH1 expression in primary tumor tissue (KM log rank p=0.148; OR=2.26, 95% CI: 0.63, 6.54, p=0.19) ; ALDH1 expression in axillary lymph node metastasis associated with poorer DFS (KM log rank p=0.037; OR=2.75, 95% CI: 0.98, 7.46, p=0.055)	No significant association with ALDH1 expression in primary tumor or axillary lymph node metastasis and DFS
Tiezzi et al[123] (2013)	Brazil	2000-2005	90	Not reported	49 (11.5)	ALDH1A1 protein expression	OS and DFS	Neoadjuvant epirubicin and docetaxel; adjuvant CMF	ALDH positive (staining of at least 5 cells in a cluster) vs ALDH negative	KM survival curves: The presence of ALDH-positive cells after neoadjuvant chemotherapy associated with worse disease-free (p=0.01) and worse overall survival (0.01)	ALDH-positive cells compared to negative: OS HR=2.54 (1.04, 6.23); Multivariable disease-free survival results not reported

Table 1 continued

Author	Study Location	Years of diagnosis	Number of participants	Length of follow-up (median(range))	Age at diagnosis (median (range) or mean (SD))	Exposure(s)	Outcomes	Chemotherapy	Unit of Analysis	Univariate results	Multivariable results
Zhou et al[124] (2013)	China	2003-2009	113	4.5 years	56.3 (33-78)	ALDH1A1 protein expression	OS and RR	None (n=22); CMF (n=11); CEF (n=52); Paclitaxel + epirubicin (n=28)	ALDH1 positive (>1% tumor cell staining) vs ALDH1 negative (≤1% cell staining)	KM survival curves, ALDH1+ associated with worse survival: Patients receiving chemotherapy without endocrine therapy: OS p=0.001, Relapse rate p=0.003; Patients receiving chemotherapy and endocrine therapy: p=0.003; No differences in relapse rate by ALDH1 positivity in patients receiving chemotherapy + endocrine therapy or endocrine therapy alone	Patients receiving chemotherapy without endocrine therapy: No association with OS after adjustment (p=0.295), Relapse rate, ALDH1+ vs ALDH1-: RR=7.493 (1.828, 3.744) p=0.005; Patients receiving chemotherapy and endocrine therapy: ALDH1+ vs ALDH1-: RR=6.759 (1.607, 28.433) p=0.009
Glutathione S-transferase Genes											
Miyake et al[125] (2012)	Japan	2004-2010	123	n/a	Not reported	GSTP1 protein expression	pCR	Paclitaxel followed by CEF (n=123)	GSTP1 positive (≥10% cells stained) vs GSTP1 negative	ER+ breast cancer: No association with pCR (OR=0.45, 95% CI: 0.11, 1.86, p=0.267); ER- breast cancer: GSTP1 expression associated with increased likelihood of pCR (OR=9.09, 95% CI: 1.65, 50.00, p=0.009)	After adjustment for menopausal status, tumor size, nodal status, tumor grade, ER, PR, and HER2 status, and Ki67expression, GSTP1 expression was associated with pCR among ER- breast cancer cases (OR=8.70, 95% CI: 1.58, 47.62, p=0.013)

Table 1 continued

Author	Study Location	Years of diagnosis	Number of participants	Length of follow-up (median(range))	Age at diagnosis (median (range) or mean (SD))	Exposure(s)	Outcomes	Chemotherapy	Unit of Analysis	Univariate results	Multivariable results
Peters et al[126] (1993)	Netherlands	1978-1987	139	Not reported	45.5	GSTA1, GSTM1, and GSTP1 protein expression	OS and DFS	Adjuvant CMF (n=139)	GSTA1 expressed vs not detectable; GSTM1 non-detectable vs low expression vs high expression; GSTP1 low vs intermediate vs high expression (tertile cut point)	KM survival curves/Univariate Cox Models: No association with disease-free survival for all GSTA1 (p=0.27), GSTM1 (p=0.24), or GSTP1 (p=0.72); No association with overall survival for all GSTs (all p-values 0.42-0.63)	No multivariable analyses reported
Arun et al[127] (2010)	USA	1997-2000	166	80 (4-163) months	57 (25-86)	GSTP1 protein expression	DFS	None (n=89); Anthracycline (n=45); Anthracycline + Taxane (n=31); CMF (n=1)	High GSTP1 expression (>70% tumor cell staining) vs low expression (≤70%)	KM survival curves, : high expression associated with better DFS: All patients p=0.09; Patients receiving chemotherapy: p=0.055	No multivariable analyses reported
Other Genes Involved in Chemotherapy Metabolism											
Horiguchi et al[128] (2002)	Japan	1985-1996	119	66 (5-126) months	51 (30-85)	DPYD protein expression	DFS and OS	None (n=32); 5-fluorouracil or derivative (n=87)	DPYD-positive (intermediate to strong tumor cell staining) vs DPD-negative (weak or no staining)	KM survival curves, All patients: Patients with DPD-positive tumors had significantly poorer DFS and OS compared to those with DPD-negative tumors ($p<0.05$); Patients receiving 5-FU chemotherapy: DPD expression significantly associated with poorer DFS ($p<0.05$)	DPD expression associated with poorer DFS (RR (se)= 0.502 (0.333) p=0.038) but not OS (RR (se)=0.629 (0.356) p=0.192)

Table 1 continued

Author	Study Location	Years of diagnosis	Number of participants	Length of follow-up (median(range))	Age at diagnosis (median (range) or mean (SD))	Exposure(s)	Outcomes	Chemotherapy	Unit of Analysis	Univariate results	Multivariable results
Yu et al[129] (2005)	China	1990-1993	197	142 (28-176) months	51 (29-76)	DPYD and TYMS protein expression	OS and DFS	Adjuvant CMF (n=197)	positive (>25% of cancer cells were stained) vs negative	KM survival curves: TS expression associated with shorter DFS (p<.0001) and OS (p<.0001); DPD expression was not associated with DFS (p=0.23) or OS (p=0.68)	Using Cox regression, TS expression: DFS OR=8.4034 (5.618, 12.5000) p<.0001; OS OR=9.1743 (5.4645, 15.3846) p<.0001
Fox et al[130] (1997)	UK	1989-1993	328	45 (5-100) months	55 (26-83)	TYMP protein expression	OS and RFS	CMF (n=127)	High TYMP expression (≥25% tumor cell staining, n=166) vs low expression (<25%, n=162))	KM survival curves, TYMP expression associated with better survival: All patients: RFS p=0.015, OS p=0.14; Patients receiving chemotherapy: RFS p=0.02, OS=0.02; No association observed in patients who did not receive chemotherapy treatment	Among node-positive patients: HR (95% CI): OS 0.4 (0.2-0.9) p=0.03, RFS 0.6 (0.3, 1.1) p=0.1; Among node-positive patients who received chemotherapy treatment: OS p=0.06; Among node-positive patients with no chemotherapy treatment: OS p=0.24
Tominaga et al[131] (2002)	Japan	1990-1992	579	8 years	Not reported; <75	TYMP protein expression	RFS and OS	Treatment group: 5'-FDUR for 6 months; Control group: no chemotherapy	TYMP scale (-, ±, +, ++): coded as 1, 2, 3, 4, a composite score including all 3 pathologist's scores was used for analysis	KM survival curves: RFS and OS more favorable in patients with higher TYMP expression scores in both the treatment and control groups (RFS: 5'-DFUR p=0.094; OS: 5'-DFUR p=0.050, Control p=0.108)	TP score independently associated with RFS in 5'-DFUR group (RR=0.856 (0.755, 0.972), p=0.016) but not in the surgery only group (p=0.35)

Table 1 continued

Author	Study Location	Years of diagnosis	Number of participants	Length of follow-up (median(range))	Age at diagnosis (median (range) or mean (SD))	Exposure(s)	Outcomes	Chemotherapy	Unit of Analysis	Univariate results	Multivariable results
Yang et al[132] (2002)	Italy	1984-1991	182	78 (3-177) months	Not Reported	TYMP protein expression	OS and DFS	Adjuvant CMF (n=51)	TYMP positive (≥50% of tumor cells showed similar or stronger staining compared to normal epithelium) vs negative	KM survival curves: TYMP expression associated with better DFS (p=0.0038) and OS (p=0.0070)	TP expression not associated with DFS or OS after adjustment
Aki et al[133] (2010)	Japan	1988-2006	217	7.6 years	53 (24-83)	TYMS, DPYD, TYMP, and OPRT mRNA expression	RFS	Adjuvant oral 5-FU (n=147)	mRNA expression level computed from its ratio to the expression of β-actin	High TS expression associated with worse survival: HR=6.67, p<0.01); High DPD expression associated with better survival: HR=0.66, p=NS; High TYMP expression: HR=1.78, p=NS; High OPRT expression: HR=3.56, p=NS compared to low expression	High TS expression: HR=10.9, p<0.01

Abbreviations: Immunohistochemical, IHC; Disease-Free Survival, DFS; Overall Survival, OS; Relapse-Free Survival, RFS; Relapse Rate, RR; Clinical Response, CR; Pathologic Complete Response, pCR

Results from Systematic Review

Cytochrome P450s

Three studies were identified which assessed the relationship between expression of cytochrome P450 genes and breast cancer outcomes, defined as survival or response to treatment.

CYP1B1 is involved in the metabolism of taxanes (e.g. paclitaxel and docetaxel). In a study of 72 breast cancer patients, CYP1B1 gene expression was associated with an increased likelihood of clinical response to chemotherapy treatment with docetaxel (unadjusted OR=1.70, 95% CI: 1.02, 2.95, p=0.0421).[117] Tissue used to measure gene expression in this study was collected prior to chemotherapy treatment. In total, 192 genes were evaluated in this study and 14 were significantly associated with clinical response (p<0.05); CYP1B1 was the only metabolizing gene which was significantly associated with clinical response. No adjustments were made for multiple comparisons. No multivariable analyses were reported.

CYP3A4 is involved in the metabolism of several chemotherapy drugs. It is involved in the deactivation of taxanes and the activation of cyclophosphamide. In a study of 89 breast cancer patients, CYP3A4 gene expression was associated with an increased likelihood of pCR (OR=1.82, 95% CI: 0.94, 3.52, p=0.0462).[119] Gene expression was measured in tumor tissue prior to chemotherapy treatment. All patients were treated with doxorubicin and paclitaxel prior to surgery and CMF (cyclophosphamide, methotrexate, and fluorouracil) after surgery. No multivariable analyses were reported for CYP3A4.

In a study of 42 breast cancer patients, CYP2D6 gene expression was significantly associated with increased likelihood of pCR (ANOVA: F=5.797, p=0.0063).[134] Gene expression was measured in tissue collected prior to chemotherapy treatment. Patients were treated with AT (doxorubicin and docetaxel) (n=29) or AC (doxorubicin and cyclophosphamide) (n=13). No multivariable analyses were reported.

Aldehyde Dehydrogenases

ALDH1A1 is involved in the metabolism of cyclophosphamide, specifically the deactivation of the active metabolite. In a study of 89 breast cancer patients, ALDH1A1 gene expression was associated with an increased likelihood of pCR (OR=2.12, 95% CI: 0.93, 4.81, p=0.0415).[119] Gene expression was measured in tumor tissue prior to chemotherapy treatment. All patients were treated with doxorubicin and paclitaxel prior to surgery and CMF (cyclophosphamide, methotrexate, and fluorouracil) after surgery. No multivariable analyses were reported for ALDH1A1.

ALDH1A1 expression was shown to be significantly associated with worse overall survival in 3 studies.[123,135,136] One of these studies also found a significant association between ALDH1A1 expression and worse disease-free survival;[123] however, two studies found no association with disease-free survival.[122,135] Additionally, one study looked at relapse rate and found a significant association with ALDH1A1 expression after adjustment for expression of several other genes (see Table 1) among patients treated with chemotherapy with or without endocrine therapy.[136] The definition of ALDH1A1 positivity differed among these studies making comparability difficult (see Table 1). Another study found that ALDH1A1 tumor gene expression was associated with better DFS and OS among TNBC patients in China and the

United States; the association persisted after adjustment for age at diagnosis, TNM stage, radiotherapy and chemotherapy treatment, and basal-like breast cancer subtype.[121] No further analyses were done by type of chemotherapy received.

No studies were identified which investigated the effect of tumor tissue gene expression levels of ALDH3A1.

Glutathione-S-Transferases

Tumor tissue gene expression of GSTP1 has been shown to be associated with worse pCR in ER-negative breast cancer patients.[125] In that study, the investigators examined GSTP1 expression in tumors in 123 Japanese breast cancer patients and its association with response to chemotherapy treatment.[125] Prior to surgery, these patients were treated with paclitaxel, followed by a combination of 5-fluorouracil, epirubicin, and cyclophosphamide. The study results suggest that GSTP1 expression may predict response to chemotherapy in ER-negative breast cancer patients, but not in ER-positive breast cancer patients.[125]

In a study of 166 breast cancer patients, high GSTP1 protein expression was marginally associated with better disease-free survival (unadjusted p=0.09)[127]. When restricted to only patients who underwent chemotherapy treatment (with anthracycline + taxane (n=31) or anthracycline alone (n=45)), the association was strengthened (p=0.055). When stratified by chemotherapy drug, GSTP1 expression was associated with marginally better disease-free survival among those who took a taxane (p=0.06) but not those who took anthracycline alone. An earlier study reported no association between disease-free or overall survival and protein expression of GSTP1, GSTA1, or GSTM1 in a population of 139 breast cancer patients with

tumor samples collected prior to chemotherapy treatment with CMF (cyclophosphamide, methotrexate, and fluorouracil);[126] however, protein expression detection methods differed between the studies. No multivariable analyses were reported for either study.

No studies were identified which investigated the effect of tumor tissue gene expression levels of GSTA1.

Other

Several genes involved in fluoropyrimidine (e.g. 5-fluorouracil, capecitabine) metabolism have been investigated in tumor level expression studies. While the majority of these studies were done in CRC patients, several studies have also investigated breast cancer.

DPYD is the most widely studied drug with respect to fluorouracil metabolism as it accounts for 80% of deactivation of this drug in the body. In a study of 119 Japanese breast cancer patients, patients with strong tumor staining for DPYD protein expression had poorer disease-free and overall survival as compared to those with weak or no DPYD staining ($p<0.05$).[128] The association with disease-free survival persisted when restricted to patients who received 5-fluorouracil or a derivative and after additional adjustment for tumor factors (RR=0.502, standard error (se)=0.333, $p=0.038$). However, another study in 197 Chinese breast cancer patients did not observe an association between DPYD protein expression and disease-free or overall survival.[129] The latter study did find an association between TYMS protein expression and disease-free and overall survival ($p<.0001$). This association persisted after adjustment for tumor characteristics (DFS: OR=8.40, 95% CI: 5.62, 12.50; OS: OR=9.17, 95% CI: 5.46, 15.38), although the magnitude of the ORs are unusually high. Another study, which

looked at mRNA expression of DPYD, TYMS, TYMP, and OPRT (also known as UMPS) in 217 Japanese breast cancer patients, found that high TYMS expression was significantly associated with poorer relapse-free survival ($p<0.01$).[133] This association persisted after adjustment for tumor characteristics (HR=10.9, $p<0.01$). This study did not find an association between DPYD, TYMP, or OPRT and relapse-free survival.

Three other studies also assessed the association between TYMP and breast cancer outcomes. In a study of 328 ER+ breast cancer patients, high TYMP expression was significantly associated with relapse-free survival ($p=0.015$), but not overall survival ($p=0.14$).[130] When the analyses were restricted to only those patients who received chemotherapy treatment (CMF; cyclophosphamide, methotrexate, and 5-fluorouracil), TYMP protein expression was significantly associated with better disease-free survival and overall survival, while no association was observed among those who did not receive chemotherapy treatment ($p=0.06$ vs $p=0.24$). After adjustment for patient and tumor characteristics, the association remained. Stratification for lymph node status showed that the relationship was only observed in node-positive breast cancer patients who were treated with chemotherapy (OS: HR=0.4, 95% CI: 0.2, 0.9). Similarly, another study in 182 breast cancer patients (both ER+ and ER-) found an association between high TYMP protein expression and better disease-free survival and overall survival; however, adjustment for tumor characteristics attenuated the observed association.[132] When stratified by chemotherapy treatment (CMF), TYMP expression was associated with better disease-free survival and overall survival among those who underwent chemotherapy (DFS: $p=0.01$; OS: $p=0.009$). A randomized trial of chemotherapy treatment with a fluorouracil derivative versus surgery only found that high TYMP protein expression was associated with

better relapse-free survival and overall survival among both groups.[131] After adjustment for patient and tumor characteristics, high TYMP expression was only associated with relapse-free survival among those treated with chemotherapy (RR=0.856, 95% CI: 0.755, 0.972, p=0.016).

Need for Consideration of Molecular Subtypes in Studies of Drug Metabolizing Enzymes and Breast Cancer Prognosis

The majority of previous studies which have examined the effect of genetic variation on chemotherapy response were done in CRC patients or in a variety of cancer types. Among those which utilized breast cancer populations, almost all of them treated breast cancer as a single disease and did not account for the heterogeneity of breast cancer. It is widely accepted that breast cancer is a heterogeneous disease and the different subtypes respond differently clinically to chemotherapy agents.[9,137] These differences in response likely stem from genetic variation in these subtypes of breast cancer.

Breast cancer is generally clinically differentiated by the expression of ER, PR, and HER2 receptors; however, evidence from research studies that measured multiple gene expression suggests that there may be more molecular subtypes based on gene expression profiling and that these subtypes may be useful in providing further prognostic information.[138-140] An algorithm was developed which differentiates breast cancer subtypes based on the expression of 50 genes, resulting in five intrinsic subtypes—basal-like, HER2-enriched, luminal A, luminal B, and normal; it was named the Prediction Analysis of Microarray 50 (PAM50).[137] The ability to further differentiate tumors on the basis of gene expression has been correlated with breast cancer outcomes, namely disease-free progression and overall survival.[141] PAM50 intrinsic

subtypes have also been shown to improve prognostic prediction compared to established clinical predictors such as IHC-based markers.[142]

Due to many similarities between TNBC and basal-like breast cancer (BLBC) tumors, such as aggressive tumor growth[138,142] and poor clinical outcomes,[9,11,143] TNBC tumors were previously considered to be overlapping with this subtype and were treated as such in many studies.[3,144] In a previous study using gene expression profiling, about 71% TNBC tumors (123 samples out of 172) were also classified as BLBC by gene expression profiling[145] and subsequent studies substantiated these findings.[146,147] Furthermore, it has been suggested that differences in chemotherapy response may be more heterogeneous in TNBC tumors than BLBC tumors.[148]

CHAPTER II

SPECIFIC AIMS

Specific Aims/Hypotheses

Aim 1: To systematically review the epidemiologic evidence available on the role of gene expression of chemotherapy metabolizing genes and breast cancer survival and response to chemotherapy.

We performed a systematic literature review in order to summarize the current state of the epidemiologic literature on the role of gene expression of chemotherapy metabolizing genes and breast cancer survival.

Aim 2: To investigate known and potentially functional genetic variants in genes known to metabolize cyclophosphamide and 5-fluorouracil for their association with disease-free survival and overall survival among all breast cancer subtypes.

We hypothesized that known and potentially functional genetic variants in genes known to metabolize cyclophosphamide and 5-fluorouracil were associated with disease-free and overall survival.

Aim 3: To investigate whether the tumor tissue expression levels of genes known to metabolize particular chemotherapy drugs are associated with disease-free and overall survival among TNBC patients.

We hypothesized that tumor tissue expression level of genes known to metabolize particular chemotherapy drugs, by influencing intracellular (or tissue) exposure dose of the drug, were associated with disease-free survival and overall survival.

Aim 4: To evaluate the correlation between germline polymorphisms and gene expression level in tumor tissue as well as the joint effect of the SNP-based gene metabolizing score and the gene expression level-based gene metabolizing score on disease-free and overall survival.

We hypothesized that the joint effect of the SNP metabolizing score, which may influence the circulating exposure of medication, and gene expression level score in tumor tissue, which may affect drug exposure at the local level, was associated with decreased disease-free and overall survival.

CHAPTER III

METHODS

Study Overview

The Shanghai Breast Cancer Survival Study (SBCSS), an established longitudinal, population-based cohort study with ongoing follow-up of outcomes, was used for the current study.[149] In brief, the cohort consists of 5,042 breast cancer survivors diagnosed with incident breast cancer between March 2002 and April 2006 and identified through the Shanghai Cancer Registry. Participants, ages 20 to 75 years, were permanent residents of Shanghai, China and recruited to the study approximately 6 months after diagnosis of primary breast cancer (range: 5.1-9.1 months). Cases were diagnosed with stage 0-IV breast cancer (based on AJCC, 6th edition); a combination of medical record review and central review of pathology slides was used to confirm breast cancer diagnoses.

Data Collection and Biological Samples

Baseline: For eligible women, information on demographic and lifestyle variables, as well as clinical and treatment variables, was collected by trained interviewers using structured questionnaires approximately 6 months following diagnosis (range 5.1-9.1 months). Information collected included demographic information, reproductive history, lifestyle factors (including smoking status, physical activity, and body mass index), and medical history (including comorbidities, usual dietary intakes, and vitamin, supplement, and alternative medicine use). Additionally, comprehensive information on cancer diagnosis and treatment, including chemotherapy, radiotherapy, Tamoxifen use, and surgery, and tumor characteristics, including

stage, grade, histology, and hormone receptor status, was obtained. Clinical data was verified through medical chart review. Tumor sections (9 5-um and 1 15-um) were obtained for 4,036 participants (80%) from the referring hospital and stored covered in paraffin at -4°C in a vacuum chamber. Exfoliated buccal cells and saliva samples were collected as the genomic DNA source for 98% of the participants. This data has been used for several studies investigating breast cancer risk and survival and genetic variants.[114,150-152]

Follow-up: Additional in-person follow-up surveys occurred at 18, 36, and 60 months post-diagnosis. Information on disease progression, recurrence, and survival was collected, as well as information on treatment, including chemotherapy. The date of last in-person contact or December 2013 (6-months prior to date of latest record linkage), whichever was more recent, was used as the censor date for event-free subjects.

Study components relevant to the specific aims of this proposal are described in further detail below.

Study Design

Using resources available in the SBCSS cohort, we investigated individual genetic variants in drug metabolizing genes among all breast cancer patients with available data (n=3,740) and tumor tissue expression levels of drug metabolizing genes in TNBC patients (n=469) and their association with disease-free and overall survival. Additionally, we evaluated the joint effect of tumor expression levels and individual genetic variants in drug metabolizing genes on disease-free and overall survival in TNBC patients with genotyping and gene

expression data available (n=312). Details on populations and specific methodologies utilized for each aim in this study are described below.

Participant Population for Aim 2

Aim 2 will be conducted among all SBCSS participants, regardless of breast cancer subtype, with available data for the SNPs of interest (see Table 2).

Participant Population for Aim 3

Of the 469 TNBC participants from the SBCSS study with available gene expression data, we excluded those whose tissue samples were collected after chemotherapy treatment (n=34) or those where the information of timing of chemotherapy treatment in relation to tissue collection was unknown (n=17). This resulted in 418 TNBC SBCSS participants being included in the analyses for Aim 3.

Participant Population for Aim 4

Participants with gene expression data available from Aim 3, who also have germline genetic information available from Aim 2, will be used in the analyses for Aim 4 (n=312). Not all overlapping cases have germline genetic information available for every SNP of interest (see Table 2).

Gene Expression Measurement

The methods for gene expression profiling were part of a large gene expression effort previously described by Baglia *et al.*[153] Briefly, using NanoString technology, gene expression levels for 311 selected genes were measured from formalin-fixed paraffin-embedded (FFPE)

breast cancer tumor tissues.[154] Using methods developed by Parker et al, breast cancer tumors were classified into subtypes: Basal-like, Luminal A, Luminal B, HER2-enriched or Normal-like breast cancer based on PAM50 genes.[137] For analysis, gene expression values were log base 2 transformed to account for non-normal distribution of the data. From the SBCSS cohort, only those patients with TNBC were included in the gene expression profiling and were included in the associated aims in the current study. More information on exclusion criteria, quality control, and normalization of the samples can be found in the Baglia *et al* paper.[153]

Selection of SNPs

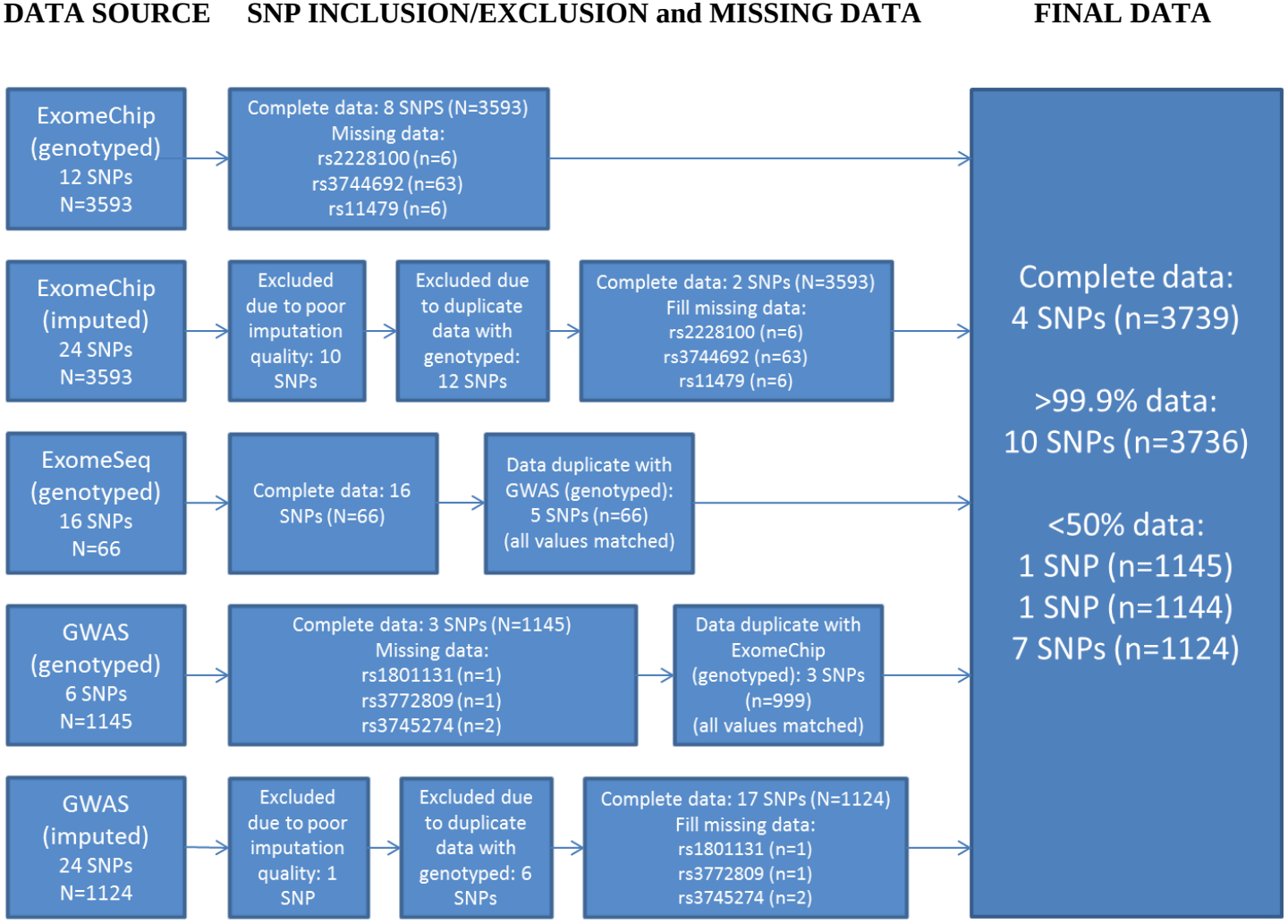
Relevant SNPs in the drug metabolizing pathways of cyclophosphamide and 5-fluorouracil were selected through two methods. First, the Pharmacogenomics Knowledgebase (PharmGKB) database[24] (https://www.pharmgkb.org/) was searched to identify SNPs in the drug metabolizing genes of interest which were previously reported as being associated with cancer outcomes (toxicity, adverse side effects, response, and survival) or were previously shown in healthy populations to alter metabolic rate. Additionally, a literature search in PubMed was conducted to find additional studies. Secondly, the University of California, Santa Cruz (UCSC) Genome Browser[155] (http://genome.ucsc.edu/cgi-bin/hgGateway) and the Exome Aggregation Consortium (ExAC) Browser[156] (http://exac.broadinstitute.org/) were both searched to identify all SNPs in the genes of interest. SNPs which were identified as functional, either nonsense, missense, frameshift, or splicing variants, were considered for inclusion in the current study. Intron variants were only included if previous research showed a correlation with cancer outcomes. The UCSC Genome Browser[155] and Ensembl Variant Effect Predictor (VEP)[157] (http://www.ensembl.org/info/docs/tools/vep/index.html) were used to predict the functional

consequence of each SNP. Variants were excluded if the minor allele frequency (MAF) in Asians was <5%, resulting in 16 SNPs in cyclophosphamide metabolizing genes and 18 SNPs in 5-fluorouracil metabolizing genes being excluded. Rare variants were not included due to reduced power to detect an association and low potential utility as predictive markers. Using HaploReg v3, SNPs were evaluated for linkage disequilibrium (LD) with other remaining SNPs identified in the previous step. The 1000 Genomes Phase 1 Asian population was used to calculate LD with a R^2 threshold of 0.5. In order to avoid redundancy, only one SNP was included for those in high LD. For SNPs in strong LD (>0.8) where only one SNP was functional, the functional SNP was chosen for analysis; this resulted in the exclusion of 2 DPYD SNPs (rs17116806 which was in high LD with rs1801159 (R^2=0.96) and rs4970722 which was in high LD with rs1801265 (R^2=1)) and 2 UMPS SNPs (rs2291078 and rs3772810) which were both in high LD with rs3772809 (R^2=0.84 and R^2=1, respectively)). In the GSTM1 gene, two missense SNPs were in strong LD (rs202002774 and rs199816990); the SNP with the higher Asian MAF (0.17 vs 0.16) was chosen for genotyping (rs202002774). There were 15 SNPs in genes involved in cyclophosphamide metabolism and 14 SNPs in genes involved in 5-fluorouracil metabolism which met all criteria.

Existing germline genetic data for members of the SBCSS cohort was collected from several previous studies and combined for the SNPs identified in genes involved in metabolism of cyclophosphamide and 5-fluorouracil (Figure 1).[158,159] Imputation was performed using established methods which have been published previously.[160,161] Briefly, genotypes were imputed for SNPs using the program MACH 1.0 and dosage data was used for statistical analyses. Quality control measures were performed prior to the statistical analyses. Both

genotyping and imputation data was available from previous studies; when available, genotyping data was used. Imputed SNP data with poor imputation scores, i.e. R^2 values less than 0.5, were removed from further analysis; although a R^2 threshold of 0.5 was chosen, all imputation scores for included data were ≥0.6. After removing duplicate imputation data and poor quality imputation data, the allele frequencies for each available SNP were calculated by dataset and compared with the published allele frequencies in Asian populations from the 1000 Genomes Phase 1 study.

Figure 1: Pre-Existing Data Available for SNPs Identified for Inclusion in the Current Study in the SBCSS and Final Sample Size After Quality Control



After quality control measures were taken, data was available for 11 SNPs in genes involved in cyclophosphamide metabolism and 12 SNPs in genes involved in 5-fluorouracil metabolism (Table 2).

Table 2: Available Data on Single Nucleotide Polymorphisms Investigated in Current Study

Chemotherapy Drug	Gene	rs ID	TNBC (N)	All Breast Cancer (N)
Cyclophosphamide	CYP2B6	rs3745274	107	1144
	CYP2C19	rs4244285	341	3736
	CYP2C19	rs4986893	341	3736
	CYP2C8	rs2071426	341	3736
	ALDH1A1	rs3764435	105	1124
	ALDH1A1	rs63319	105	1124
	ALDH3A1	rs2228100	342	3739
	ALDH3A1	rs887241	342	3739
	ALDH3A1	rs3744692	341	3736
	GSTA1	rs3957357	105	1124
	GSTP1	rs1695	341	3736
5-Fluorouracil	DPYD	rs17376848	105	1124
	DPYD	rs1801159	341	3736
	DPYD	rs1801265	341	3736
	DPYD	rs72728438	105	1124
	MTHFR	rs1801131	342	3739
	MTHFR	rs1801133	105	1124
	MTHFR	rs2274976	341	3736
	TYMP	rs11479	341	3736
	TYMS	rs2847153	107	1145
	TYMS	rs2853533	105	1124
	UMPS	rs1801019	341	3736
	UMPS	rs3772809	342	3739

Chemotherapy Information Collection

Information on chemotherapy administration was collected at the baseline interview via patient report. The most commonly prescribed chemotherapy drugs and associated sample sizes can be found in Table 3. In addition to the chemotherapy drug type, information on the dates that each specific therapy was started and stopped, the dose, the number of cycles of the therapy, and the total duration of treatment was collected for each chemotherapy drug administered to the patient. Due to this granular data collection, we are able to account for concurrent chemotherapies as well as duration of treatment. Of the 469 TNBC participants, 418 tumor tissue samples were collected prior to chemotherapy treatment, including 28 who did not undergo chemotherapy treatment. For 17 participants, the timing of chemotherapy could not be determined because information was only collected for month and year and treatment, not day; for these samples, surgery and chemotherapy treatment occurred in the same month. Aim 3, which evaluates tumor level gene expression data, will only include those participants whose tissues samples were known to be collected prior to chemotherapy treatment because it has been shown that chemotherapy can alter tumor tissue gene expression.[162,163]

Table 3: Chemotherapy Use in the Current Study

	All Breast Cancer (Aim 2)	TNBC (Aim 3)
Chemotherapy	**N(%)**	**N(%)**
Any	3397 (90.9)	441 (94.0)
Neoadjuvant	181 (5.3)	34 (7.7)
Adjuvant	1528 (45.0)	390 (88.4)
Unknown	1688 (49.7)	17 (3.9)
Cyclophosphamide	2594 (69.4)	337 (71.9)
Neoadjuvant	97 (3.7)	26 (7.7)
Adjuvant	1167 (45.0)	296 (87.8)
Unknown	1330 (51.3)	15 (4.5)
5-Fluorouracil	2626 (70.2)	354 (75.5)
Neoadjuvant	93 (3.5)	25 (7.1)
Adjuvant	1183 (45.1)	316 (89.3)
Unknown	1350 (51.4)	13 (3.7)
Other Chemotherapy Drugs		
Anthracyclines[1]	2078 (55.6)	264 (56.3)
Taxanes[2]	264 (7.1)	36 (7.7)

Note: Information on timing of chemotherapy not relevant for Aim 2 study objectives

[1]Including doxorubicin and epirubicin

[2]Including paclitaxel and docetaxel

Outcome Ascertainment

The primary outcomes of interest are disease-free survival and overall survival which were calculated from the date of initial breast cancer diagnosis for all patients. Participants were followed-up to obtain information on survival status and cancer recurrence at 18, 36, and 60 months after breast cancer diagnosis. The follow-up rates for the in-person interviews are 92.8%, 86.4%, and 78.9% for the 18, 36, and 60 month interviews, respectively. Survival information was obtained using annual record linkage with the Shanghai Vital Statistics Registry for all

participants, including those lost to follow-up. Observation time was censored at the time of the event of interest (recurrence or death) or, for those who did not experience an event, date of last follow-up, the later of either last in-person survey or annual record linkage. For the disease-free analysis, individuals who died of non-breast cancer-related causes were censored at date of death.

General Analytic Approach

In this study, we examined expression levels of drug metabolizing genes in the tumor tissue and germline genetic variants within the individual to better understand how these factors individually and jointly influence variability in prognosis following treatment. We focused on cyclophosphamide and 5-fluorouracil since they were the two most commonly used chemotherapy drugs in the SBCSS. In the United States, these two drugs, in combination with other chemotherapy agents, are among the most commonly used chemotherapy drugs to treat breast cancer.[164]

The primary outcomes of interest for the current study were disease-free survival (DFS) and overall survival (OS). Kaplan-Meier survival curves were generated to evaluate the univariate association between exposures of interest and DFS and OS. The log-rank *p*-value was used to test significance.

Using linear models for continuous variables and χ^2 contingency tables for categorical variables, the association between selected demographic, lifestyle, treatment, and clinical variables was characterized prior to multivariable modeling.

Cox models were used to estimate the associated hazards ratios (HRs) and to calculate resulting 95% confidence intervals and associated p-values.[165] The proportional hazards assumption was formally tested through the inclusion of an interaction term between exposures and time since diagnosis; no interactions with survival were observed. Entry time was defined as date of diagnosis and exit time was defined as date of event (either recurrence or death) or date of last follow-up/record linkage. For variables treated as categorical, *p*-values for trends were calculated by treating the variable as a continuous variable. Analyses were stratified by various factors, including whether the chemotherapy drug of interest was taken and timing of events (defined as early events (<3 years since diagnosis) and later events (≥3 years post diagnosis), to evaluate potential effect measure modifiers; likelihood ratio tests were used to test potential effect measure modification.

Information on toxicity was not available for the SBCSS cohort. However, due to the comprehensive collection of chemotherapy use, including number of cycles of chemotherapy received for each chemotherapy drug, number of cycles of drug taken was used as a surrogate for toxicity. Logistic regression was used to assess whether the outcomes of interest were associated with number of cycles (≥6 cycles (median) vs <6 cycles). The completion of 6 cycles was chosen because a typical chemotherapy regimen including cyclophosphamide or 5-fluorouracil is 6 cycles.[166] This was further corroborated in our data where 74% of patients treated with cyclophosphamide completed 6 cycles and 73% of patients treated with 5-fluorouracil completed 6 cycles.

More detailed specific statistical analyses for each aim are described in the chapter pertaining to that aim.

CHAPTER IV

POLYMORPHISMS IN CHEMOTHERAPY METABOLIZING GENES AND BREAST CANCER OUTCOMES

Aim 2-Specific Methods

Based on knowledge of cyclophosphamide and 5-fluorouracil metabolism and previous research which found associations between polymorphisms of interest and cancer outcomes (toxicity, side effects, dose delay, survival, and drug clearance), the predicted effect of each SNP was hypothesized (Table 4 for cylcophosphamide and Table 5 for 5-fluorouracil).

Table 4: Hypothesized Effects of SNPs in Genes Involved in Cyclophosphamide Metabolism

Gene	rs ID	Functional Consequence[1]	Phenotype	Summation of Evidence[2]	Predicted Effect on Activity[3]
Activation					
CYP2B6					
	rs3745274	Missense	GG	Increased likelihood of dose reduction, increased risk of adverse side effects	Increased activity
			GT	Increased likelihood of dose reduction, decreased risk of adverse side effects	Slightly decreased activity
			TT	Decreased likelihood of dose reduction, decreased risk of adverse side effects	Decreased activity
CYP2C19					
	rs4244285	Synonymous	GG	Increased risk of ovarian toxicity, decreased survival	Increased activity
			AA + AG	Decreased risk of ovarian toxicity, increased survival	Decreased activity
	rs4986893	Nonsense	Allele G	No evidence in cyclophosphamide; Other drugs: Associated with increased metabolism to active metabolite and increased response	Increased activity
			Allele A	No evidence in cyclophosphamide; Other drugs: Associated with decreased metabolism to active metabolite and decreased response	Decreased activity
CYP2C8					
	rs2071426	Splice Acceptor		No published evidence found	Unknown
Deactivation					
ALDH1A1					
	rs3764435	Intronic	Allele A	Increased risk of toxicity	Decreased activity
			Allele C	Decreased risk of toxicity	Increased activity
	rs63319	Intronic	GG	No association with toxicity	Unknown (No association?)
			GG + TT	No association with toxicity	
ALDH3A1					
	rs2228100	Missense	GG	Decreased likelihood of cystitis	Increased activity
			CC + CG	Increased likelihood of cystitis	Decreased activity
	rs887241	Missense		No published evidence found	Unknown
	rs3744692	Missense		No published evidence found	Unknown
GSTP1					
	rs1695	Missense	AA	Increased response, decreased severity of toxicity, increased progression-free survival	Decreased activity
			AG	Increased response, decreased severity of toxicity, decreased progression-free survival	Slightly decreased activity
			GG	Decreased response, increased severity of toxicity, decreased progression-free survival	Increased activity
GSTA1					
	rs3957357	Upstream variant 2KB	AA	Increased risk of anemia	Decreased activity
			AG	Increased risk of anemia	Slightly decreased activity
			GG	Decreased risk of anemia	Increased activity

[1]Source: HaploReg v3
[2]Source: PharmGKB
[3]Based on results from previous studies and role in chemotherapy metabolism pathway

Table 5: Hypothesized Effects of SNPs in Genes Involved in 5-Fluorouracil Metabolism

Gene	rs ID	Functional Consequence[1]	Phenotype	Summation of Evidence[2]	Predicted Effect on Activity[3]
Activation					
TYMP					
	rs11479	Missense	GG	Decreased risk of drug toxicity	Decreased activity
			GA	Increased risk of drug toxicity	Increased activity
			AA	Increased risk of drug toxicity	Increased activity
UMPS					
	rs1801019	Missense	GG	Decreased risk of toxicity	Decreased activity
			GC	Decreased risk of toxicity	Slightly increased activity?
			CC	Increased risk of toxicity	Increased activity
	rs3772809	Missense	AA	Improved response	Increased activity
			AG	Intermediate response	Slightly increased activity
			GG	Worse response	Decreased activity
Deactivation					
DPYD					
	rs17376848	Synonymous	AA + AG	Increased risk of toxicity	Decreased activity
			GG	Decreased risk of toxicity	Increased activity
	rs1801265	Missense	GG + GA	Increased risk of adverse side effects, no association with toxicity	Decreased activity
			AA	Decreased risk of adverse side effects, no association with toxicity	Increased activity
	rs1801159	Missense	TT + CT	Decreased risk of adverse side effects, no association with response	Increased activity
			CC	Increased risk of adverse side effects, no association with response	Decreased activity
	rs72728438	Intronic	TT	Normal DPYD activity	Increased activity
			CC + CT	Decreased DPYD activity	Decreased activity
Response					
MTHFR					
	rs1801131	Missense	TT	Reduced response to treatment	Increased activity
			GG + GT	Better response to treatment	Decreased activity
	rs1801133	Missense	GG + GA	Decreased risk of toxicity	Increased activity
			AA	Increased risk of toxicity	Decreased activity
	rs2274976	Missense		No published evidence found	Unknown
TYMS					
	rs2847153	Intronic	GG + GA	Decreased likelihood of survival	Increased expression
			AA	Increased likelihood of survival	Decreased expression
	rs2853533	Missense		No published evidence found	Unknown

[1]Source: HaploReg v3
[2]Source: PharmGKB
[3]Based on results from previous studies and role in chemotherapy metabolism pathway

In addition to findings from previous studies, Gene-Tissue Expression (GTEx) expression quantitative trait loci (eQTL) data in normal breast tissue was used to inform the effect that the selected SNPs had on gene expression of the genes of interest. The correlation between SNPs of interest and genes of interest is in Table 6.

Table 6: Gene-Tissue Expression (GTEx) eQTL Results

Gene	SNP rs ID	Chr:Position[1]	Reference Allele	Effect Allele	Beta	*p*-value
ALDH1A1	rs3764435	9:75516876	A	C	-0.024	0.70
ALDH1A1	rs63319	9:75524784	G	T	0.004	0.95
ALDH3A1	rs2228100	17:19642952	G	C	-0.248	0.01
ALDH3A1	rs887241	17:19645938	A	C	-0.104	0.29
CYP2C8	rs4244285	10:96541616	G	A	0.341	0.02
CYP2C8	rs2071426	10:96828323	T	C	0.010	0.93
DPYD	rs72728438	1:97847874	T	C	-0.195	0.01
DPYD	rs17376848	1:97915624	A	G	-0.091	0.58
DPYD	rs1801159	1:97981395	T	C	0.094	0.25
GSTP1	rs1695	11:67352689	A	G	-0.361	<.0001
MTHFR	rs2274976	1:11850927	C	T	0.145	0.38
MTHFR	rs1801131	1:11854476	T	G	0.061	0.37
MTHFR	rs1801133	1:11856378	G	A	-0.044	0.55
TYMP	rs11479	22:50964236	G	A	-0.220	0.04
TYMS	rs2853533	18:658064	G	C	0.217	0.09
TYMS	rs2847153	18:661647	G	A	0.121	0.28
UMPS	rs1801019	3:124456742	G	C	-0.032	0.61

[1]Build 37

We created gene scores using an additive model to incorporate all of the SNPs into a single score. This approached has been used previously in order to account for the effects of multiple alleles on outcomes of interest.[167]

For cyclophosphamide, three gene scores were created based on *a priori* hypotheses as to the effect of each SNP in genes involved in cyclophosphamide metabolism and the association with disease-free and overall survival was evaluated for each score. The first score included SNPs in genes which are involved in the activation of cyclophosphamide into the active metabolite (CYP genes). Alleles in the SNP associated with faster metabolism were coded as 1; therefore an individual with two such alleles would be given a score of 2 while heterozygotes would get a score of 1. Individuals who do not carry the allele associated with faster metabolism would be given a score of 0. The scores from the four SNPs identified for study in the CYP genes were then summed to create the cyclophosphamide activation gene score. The possible range of this score was 0-8; in our data, the range was 2.67-8.0 with a median value of 6.0. We would expect that individuals with higher scores would have higher circulating levels of the active metabolite and would have a longer exposure to cyclophosphamide as compared with those with lower scores (HR<1.0).

Similarly, SNPs in genes involved in the deactivation, or clearance, of the active metabolite of cyclophosphamide (ALDHs and GSTs) were coded so that faster metabolizers were given higher scores. The scores from the seven SNPs identified for study in the ALDH and GST genes were then summed to create the cyclophosphamide deactivation gene score. The possible range of this score was 0-14; in our data, the range was 3.1-14.0 with a median value of 8.75. We would expect that those with higher scores would have a shorter exposure period to the active metabolite and would have poorer survival rates as compared with those with lower scores (HR>1.0).

A total score was calculated by combining the activation and deactivation scores. Coding was done such that individuals with higher scores had longer exposure to the active metabolite of cyclophosphamide; that is, more SNPs associated with faster activation and fewer SNPs associated with faster clearance. The possible range of this score was 0-22; in our data, the range was 2.67-18.84 with a median value of 11.9. We would expect that those with higher scores would have a longer exposure period to the active metabolite and would have better survival as compared to those with lower scores (HR<1.0).

For 5-fluorouracil, five gene scores were created based on *a priori* hypotheses as to the effect of each SNP in genes involved in 5-fluorouracil metabolism and response and the association with disease-free and overall survival was evaluated for each score. The first score included SNPs in genes which are involved in the activation of 5-fluorouracil into the active metabolite (TYMP and UMPS genes). As previously described, 5-fluorouracil SNPs were coded using the same methods described for cyclophosphamide SNPs. The scores from the three SNPs identified for study in the TYMS and UMPS genes were then summed to create the 5-fluorouracil activation gene score. The possible range of this score was 0-6; in our data, the range was 0.0-6.0 with a median value of 2.9. We would expect that those with higher scores would have a higher level of active metabolite and would have better survival as compared to those with lower scores (HR<1.0).

Similarly, SNPs in the DPYD gene involved in the deactivation, or clearance, of the active metabolite of 5-fluorouracil were coded so that faster metabolizers were given higher scores. The scores from the four SNPs identified for study in the DPYD gene were then summed to create the 5-fluorouracil deactivation gene score. The possible range of this score was 0-8; in

our data, the range was 0.0-8.0 with a median value of 6.0. We would expect that those with higher scores would have a shorter exposure period to the active metabolite and would have poorer survival as compared with those with lower scores (HR>1.0).

A third score was calculated which included genes involved in the response of the 5-fluorouracil active metabolite in the tumor cell. SNPs in the TYMS and MTHFR genes were coded based on the hypothesis that individuals with higher scores had potentially improved treatment response compared to those with lower scores. The scores from the five SNPs identified for study in the TYMS and MTHFR genes were then summed to create the 5-fluorouracil response gene score. The possible range of this score was 0-10; in our data, the range was 2.17-8.88 with a median value of 5.0. We would expect that those with higher scores would have lower levels of TS and higher levels of folate and would have better survival as compared with those with lower scores (HR<1.0).

Two total scores were calculated for 5-fluorouracil. The first total score was calculated by combining the activation and deactivation scores. Coding was done such that individuals with higher scores had longer exposure to the active metabolite of 5-fluorouracil; that is, more SNPs associated with faster activation and fewer SNPs associated with faster clearance. The possible range of this score was 0-14; in our data, the range was 0.0-11.0 with a median value of 4.6. We would expect that those with higher scores would have a longer exposure period to the active metabolite and would have better survival as compared to those with lower scores (HR<1.0).

The second total score additionally included the TYMS and MTHFR SNPs. The reasoning for the creation of this second gene score was that TYMS and MTHFR are involved in response to 5-fluorouracil rather than metabolism; however, these genes have been studied in

conjunction with the activating and deactivating genes in previous studies in CRC patients. The possible range of this score was 0-24; in our data, the range was 5.7-16.5 with a median value of 10.6. We would expect that those with higher scores would have a longer exposure period to the active metabolite and would have better survival as compared to those with lower scores (HR<1.0).

Using Cox proportional hazards models, the associations between each of these scores, which were treated as categorical variables, and survival outcomes were evaluated. The *p*-values for trends were calculated by treating the SNP score as a continuous variable.

Results

Among the 3,739 breast cancer cases included in Aim 2 of this study, there were 516 recurrences/breast cancer-specific deaths and 465 all-cause deaths over a median follow-up of 5.3 years (range: 0.7-8.9 years). As shown in Table 7, 5-year DFS and OS rates were associated with age at diagnosis and were significantly positively correlated with education level, Tamoxifen use, estrogen and progesterone receptor positivity, and having had a mastectomy. We also found that 5-year DFS and OS rates were inversely associated with TNM stage, tumor grade, BMI, HER2 positivity, and radiotherapy. Additionally, 5-year OS was inversely associated with number of live births and menopausal status.

Table 7: Demographic and Clinical Predictors for Breast Cancer Survival in Breast Cancer Cases Included in Aim 2

		Disease-Free Survival			Overall Survival		
Characteristics	**N**	**Event, No.**	**5-Yr Survival Rate, %[1]**	***P***	**Deaths, No.**	**5-Yr Survival Rate, %[1]**	***P***
Age at diagnosis, y							
<40	176	33	79.6		28	88.5	
40-49	1464	174	87.3	0.04	146	91.6	0.001
50-59	1118	164	84.7		138	90.2	
≥60	981	145	85.0		153	88.0	
Education							
Elementary School or Less	426	88	79.3		90	84.5	
Middle School	1275	190	84.6	<.0001	169	89.2	<.0001
High or Vocational School	1421	180	86.7		156	90.8	
College or University	617	58	89.7		50	94.0	
Income (yuan/person/month)							
<500	373	62	82.5		53	88.4	
500 - <700	580	83	85.3		79	88.9	
700 - <1000	1092	168	84	0.09	151	89.1	0.13
1000 - <2000	1220	151	87.2		135	91.0	
≥2000	473	52	87.9		47	92.9	
Body Mass Index (kg/m^2)							
<25	2394	304	86.7		279	90.9	
25-29.99	1133	163	85.2	0.0003	149	89.2	0.02
≥30	213	49	75.9		37	85.3	
Menopausal Status							
Premenopausal	1812	237	86.2	0.53	194	91.2	0.002
Postmenopausal	1928	279	85.1		271	89.0	
Tamoxifen Use							
No	1802	294	82.6	<.0001	276	86.9	<.0001
Yes	1936	222	88.4		189	93.1	
TNM Stage							
0-I	1397	82	93.8		77	95.7	
IIA	1233	141	88.1	<.0001	111	93.0	<.0001
IIB	573	117	78.9		110	85.4	
III-IV	362	151	56.4		146	66.2	
Unknown	175	25	84.9		21	89.6	
Grade							
1	428	34	91.4		24	96.2	
2	1429	160	88.0	<.0001	153	92.0	<.0001
3	1049	207	79.7		193	84.9	
Unknown	834	115	85.8		95	90.3	
Estrogen Receptor Status							
+	2401	286	87.7	<.0001	246	92.3	<.0001
-	1293	216	82.3		205	86.6	
Unknown	46	14	65.6		14	71.7	
Progesterone Receptor Status							
+	2189	263	87.6	0.0002	222	92.4	<.0001
-	1497	238	83.1		228	87.2	
Unknown	54	15	71.4		15	75.9	

HER2 Status							
+	778	132	82.2		119	87.3	
Borderline	219	24	88.4	0.004	19	92.2	0.01
-	1859	236	86.8		223	91.1	
Unknown	884	124	85.3		104	89.9	
Chemotherapy							
No	342	39	87.9	0.19	42	90.5	0.98
Yes	3398	477	85.3		423	90.1	
Radiotherapy							
No	2540	272	88.8	<.0001	251	92.2	<.0001
Yes	1200	244	78.8		214	85.5	
Mastectomy							
No	236	42	80.2	0.01	38	85.4	0.04
Yes	3504	474	85.9		427	90.4	
No. of Live Births							
0	30	5	82.4		5	89.7	
1	2531	329	86.3	0.29	275	91.1	0.0002
2	627	92	85.1		88	88.6	
≥3	403	72	82.2		79	87.0	
Family History of BC							
No	3533	490	85.6	0.74	446	89.9	0.14
Yes	207	26	86.0		19	93.7	

[1]Survival rate calculated using life table analysis method

Twenty-three SNPs were investigated in the current study. The frequencies of alleles for each SNP were compared with the published allele frequencies in Asian populations. Genotyped and imputed data were checked individually (see Appendix B: Table B1for allele frequencies by study). The allele frequencies in all datasets were similar to the published values from 1000 Genomes (http://www.1000genomes.org/). To check allele frequencies of imputed data, dosage data was rounded to the nearest whole number (0, 1, or 2); however, dosage data was used for all analyses involving imputed data. After quality checks of the data, all data were combined into one dataset and the overall allele frequencies were again compared to 1000 Genomes[168] data (Table 8).

Table 8: Information on SNPs Included in Current Study

Gene	rs ID	Chromosome Position[1]	Alleles[3]	Allele Frequency in Asians[3]	Allele Frequency in Study Population[4]	Functional Consequence[2]
Cyclophosphamide						
Activation						
CYP2B6	rs3745274	chr19:41512841	G, T	G=0.82	G=0.81	Missense
CYP2C19	rs4244285	chr10:96541616	G, A	A=0.33	A=0.33	Synonymous
CYP2C19	rs4986893	chr10:96540410	G, A	G=0.95	G=0.94	Nonsense
CYP2C8	rs2071426	chr10:96828323	T, C	T=0.93	T=0.93	Splice Acceptor
Deactivation						
ALDH1A1	rs3764435	chr9:75516876	A, C	C=0.54	C=0.53	Intronic
ALDH1A1	rs63319	chr9:75524784	G, T	G=0.42	G=0.44	Intronic
ALDH3A1	rs2228100	chr17:19642952	G, C	G=0.57	G=0.58	Missense
ALDH3A1	rs887241	chr17:19645938	A, C	C=0.94	C=0.94	Missense
ALDH3A1	rs3744692	chr17:19643672	C, T	T=0.06	T=0.07	Missense
GSTA1	rs3957357	chr6:52668687	A, G	G=0.88	G=0.86	Unknown
GSTP1	rs1695	chr11:67352689	A, G	A=0.83	A=0.80	Missense
5-Fluorouracil						
Activation						
TYMP	rs11479	chr22:50964236	G, A	A=0.25	A=0.21	Missense
UMPS	rs1801019	chr3:124456742	G, C	C=0.19	C=0.18	Missense
UMPS	rs3772809	chr3:124462824	A, G	A=0.94	A=0.94	Missense
Deactivation						
DPYD	rs17376848	chr1:97915624	A, G	G=0.12	G=0.10	Synonymous
DPYD	rs1801159	chr1:97981395	T, C	T=0.74	T=0.73	Missense
DPYD	rs1801265	chr1:98348885	G, A	A=0.94	A=0.91	Missense
DPYD	rs72728438	chr1:97847874	T, C	T=0.80	T=0.77	Intronic
Response						
MTHFR	rs1801131	chr1:11854476	T, G	G=0.19	G=0.18	Missense
MTHFR	rs1801133	chr1:11856378	G, A	G=0.63	G=0.56	Missense
MTHFR	rs2274976	chr1:11850927	C, T	C=0.90	C=0.91	Missense
TYMS	rs2847153	chr18:661647	G, A	A=0.40	A=0.36	Intronic
TYMS	rs2853533	chr18:658064	G, C	G=0.46	G=0.49	Missense

[1]Build 19

[2]Source: HaploReg v3; forward strand alleles

[3]Source: 1000 Genomes, forward strand alleles

[4]Includes genotyped and imputed data

Cyclophosphamide

The association between each SNP in genes involved in metabolism of cyclophosphamide and DFS and OS among all breast cancer participants was evaluated (Table 9) where the allele associated with faster metabolism was coded as 1 and the allele associated with slower metabolism was coded as 0. After adjustment for age at breast cancer diagnosis and tumor grade, the G allele in SNP rs4986893 in the CYP2C19 gene was associated with better DFS (HR=0.76, 95% CI: 0.60, 0.97) and OS (HR=0.76, 95% CI: 0.59, 0.97); however, the association was no longer significant after adjustment for multiple comparisons (corrected p>.005). No significant associations between any other SNPs and DFS or OS were observed. Additional adjustment for education, BMI, menopausal status, tamoxifen use, ER status, PR status, HER2 status, radiotherapy, and mastectomy did not materially alter the observed associations.

When the results were stratified by whether cyclophosphamide was received, the association between SNP rs4986893 and DFS and OS was similar among those who took cyclophosphamide and those who did not (Table 10). The inverse correlation between the T allele in the SNP rs2071426 and survival appears to be limited to those who did not take cyclophosphamide (DFS: HR=0.75, 95% CI: 0.53, 1.06; OS: HR=0.67, 95% CI: 0.47, 0.97); however, the interaction was not significant.

Table 9: Associations Between Individual SNPs in Cyclophosphamide Metabolizing Genes and DFS and OS Among All Breast Cancer Cases

				All Participants							
				Disease-Free Survival				Overall Survival			
Gene	**rs ID**	**Effect Allele Frequency**	**N**	**HR (95% CI)[1]**	***P***	**HR (95% CI)[2]**	***P***	**HR (95% CI)[1]**	***P***	**HR (95% CI)[2]**	***P***
Activation[3]											
CYP2B6	rs3745274	G=0.81	1144	1.26 (0.91, 1.74)	0.16	1.26 (0.91, 1.75)	0.16	1.08 (0.77, 1.51)	0.67	1.10 (0.79, 1.55)	0.57
CYP2C19	rs4244285	A=0.33	3736	0.98 (0.86, 1.12)	0.81	0.98 (0.86, 1.12)	0.76	0.96 (0.84, 1.10)	0.56	0.94 (0.82, 1.08)	0.39
CYP2C19	rs4986893	G=0.94	3736	0.76 (0.60, 0.97)	0.03	0.74 (0.58, 0.94)	0.01	0.76 (0.59, 0.97)	0.03	0.75 (0.58, 0.96)	0.02
CYP2C8	rs2071426	T=0.93	3736	0.89 (0.70, 1.12)	0.31	0.85 (0.67, 1.08)	0.18	0.81 (0.64, 1.04)	0.10	0.79 (0.62, 1.01)	0.06
Deactivation[4]											
ALDH1A1	rs3764435	C=0.53	1124	1.11 (0.85, 1.44)	0.44	1.08 (0.83, 1.39)	0.57	1.23 (0.93, 1.63)	0.15	1.18 (0.89, 1.57)	0.25
ALDH1A1	rs63319	G=0.44	1124	1.05 (0.80, 1.37)	0.73	1.03 (0.79, 1.34)	0.82	1.06 (0.80, 1.42)	0.68	1.05 (0.78, 1.40)	0.74
ALDH3A1	rs2228100	G=0.48	3739	1.10 (0.97, 1.25)	0.13	1.14 (1.01, 1.30)	0.04	1.01 (0.88, 1.15)	0.89	1.04 (0.91, 1.19)	0.57
ALDH3A1	rs887241	C=0.94	3739	1.05 (0.81, 1.36)	0.70	1.00 (0.77, 1.29)	0.98	1.16 (0.87, 1.54)	0.32	1.09 (0.82, 1.46)	0.54
ALDH3A1	rs3744692	T=0.07	3736	1.12 (0.90, 1.41)	0.32	1.18 (0.94, 1.47)	0.16	1.16 (0.92, 1.47)	0.21	1.20 (0.95, 1.51)	0.13
GSTA1	rs3957357	G=0.86	1124	1.05 (0.75, 1.47)	0.77	1.01 (0.72, 1.42)	0.94	1.06 (0.73, 1.53)	0.78	1.02 (0.70, 1.49)	0.92
GSTP1	rs1695	A=0.80	3736	1.14 (0.97, 1.34)	0.12	1.18 (1.00, 1.38)	0.05	1.05 (0.89, 1.24)	0.57	1.07 (0.90, 1.26)	0.44

[1]Adjusted for age at diagnosis and tumor grade

[2]Adjusted for age at diagnosis, tumor grade, education, BMI, menopausal status, tamoxifen use, ER status, PR status, HER2 status, radiotherapy, and mastectomy

[3]Allele associated with faster metabolism coded as 1, HR<1 expected

[4]Allele associated with faster metabolism coded as 1, HR>1 expected

Table 10: Associations Between Individual SNPs in Cyclophosphamide Metabolizing Genes and DFS and OS Among All Breast Cancer Cases Stratified by Whether Cyclophosphamide was Taken

		Participants who Took Cyclophosphamide				Participants who Did Not Take Cyclophosphamide					
		Disease-Free Survival		Overall Survival		Disease-Free Survival		Overall Survival		DFS	OS
Gene	**rs ID**	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***	P_{int}	P_{int}
Activation[3]											
CYP2B6	rs3745274	1.38 (0.94, 2.04)	0.10	1.14 (0.76, 1.72)	0.52	0.95 (0.52, 1.75)	0.88	0.87 (0.47, 1.62)	0.66	0.27	0.39
CYP2C19	rs4244285	0.97 (0.82, 1.15)	0.69	0.98 (0.82, 1.17)	0.82	1.01 (0.82, 1.24)	0.94	0.93 (0.74, 1.15)	0.49	0.76	0.65
CYP2C19	rs4986893	0.77 (0.56, 1.05)	0.10	0.75 (0.54, 1.04)	0.09	0.78 (0.54, 1.14)	0.20	0.79 (0.53, 1.16)	0.22	0.95	0.82
CYP2C8	rs2071426	1.02 (0.74, 1.41)	0.91	0.93 (0.67, 1.31)	0.69	0.75 (0.53, 1.06)	0.10	0.67 (0.47, 0.97)	0.03	0.18	0.15
Deactivation[4]											
ALDH1A1	rs3764435	1.02 (0.75, 1.38)	0.90	1.13 (0.81, 1.56)	0.48	1.33 (0.79, 2.22)	0.28	1.59 (0.90, 2.80)	0.11	0.25	0.27
ALDH1A1	rs63319	1.04 (0.76, 1.42)	0.82	0.98 (0.70, 1.38)	0.93	0.98 (0.58, 1.66)	0.94	1.29 (0.73, 2.29)	0.38	0.94	0.36
ALDH3A1	rs2228100	1.06 (0.90, 1.24)	0.49	0.96 (0.81, 1.14)	0.63	1.17 (0.96, 1.44)	0.13	1.08 (0.88, 1.34)	0.45	0.44	0.37
ALDH3A1	rs887241	1.06 (0.75, 1.49)	0.75	1.18 (0.81, 1.72)	0.38	1.07 (0.72, 1.59)	0.73	1.14 (0.73, 1.77)	0.57	0.95	0.81
ALDH3A1	rs3744692	1.17 (0.87, 1.56)	0.29	1.20 (0.89, 1.62)	0.23	1.08 (0.75, 1.55)	0.69	1.13 (0.78, 1.63)	0.53	0.74	0.77
GSTA1	rs3957357	1.17 (0.79, 1.74)	0.44	1.16 (0.75, 1.79)	0.50	0.80 (0.43, 1.49)	0.48	0.80 (0.39, 1.62)	0.53	0.34	0.41
GSTP1	rs1695	1.10 (0.89, 1.35)	0.39	1.21 (0.93, 1.57)	0.44	1.09 (0.88, 1.36)	0.15	0.98 (0.76, 1.27)	0.88	0.55	0.56

[1]Adjusted for age at diagnosis and tumor grade

[2]Adjusted for age at diagnosis, tumor grade, education, BMI, menopausal status, tamoxifen use, ER status, PR status, HER2 status, radiotherapy, and mastectomy

[3]Allele associated with faster metabolism coded as 1, HR<1 expected

[4]Allele associated with faster metabolism coded as 1, HR>1 expected

Each of the three gene scores created for SNPs in genes involved in cyclophosphamide metabolism were evaluated for their association with DFS and OS among all breast cancer cases (Table 11). Although no significant associations were observed, the point estimate of each score was in the hypothesized direction for OS; for DFS, the point estimate was in the hypothesized direction for the total score and activation score.

When stratified by whether cyclophosphamide was taken, no significant associations were observed among the three gene scores and DFS or OS among those who underwent treatment with cyclophosphamide and those who did not (Table 11). The effect estimate for the total score was very similar between those who took cyclophosphamide and those who did not.

Table 11: Association between Cyclophosphamide Gene Scores and Breast Cancer Outcomes Among All Breast Cancer Cases and Stratified by Whether Cyclophosphamide was Taken

			Disease-Free Survival		Overall Survival	
Role in Metabolism	**# of SNPs**	**N**	**HR (95% CI)[4]**	***P***	**HR (95% CI)[4]**	***P***
All Participants						
Activation[1]	4	1141	1.01 (0.85, 1.19)	0.93	0.94 (0.78, 1.12)	0.47
Deactivation[2]	7	1124	1.07 (0.97, 1.18)	0.20	1.03 (0.92, 1.15)	0.64
Total Score[3]	11	1124	0.95 (0.88, 1.04)	0.29	0.96 (0.88, 1.06)	0.44
Participants Who Took Cyclophosphamide						
Activation[1]	4	889	0.99 (0.81, 1.21)	0.94	0.90 (0.73, 1.12)	0.36
Deactivation[2]	7	878	1.06 (0.94, 1.19)	0.36	1.01 (0.89, 1.15)	0.85
Total Score[3]	11	878	0.96 (0.87, 1.06)	0.43	0.97 (0.87, 1.08)	0.56
Participants who Did Not Take Cyclophosphamide						
Activation[1]	4	252	1.07 (0.76, 1.50)	0.71	0.99 (0.70, 1.42)	0.97
Deactivation[2]	7	246	1.12 (0.92, 1.36)	0.27	1.06 (0.86, 1.32)	0.58
Total Score[3]	11	246	0.93 (0.78, 1.10)	0.39	0.95 (0.79, 1.14)	0.57

[1]Includes SNPs from CYP genes, allele associated with faster metabolism coded as 1, HR<1 expected

[2]Includes SNPs from ALDH and GST genes, allele associated with faster metabolism coded as 1, HR>1 expected

[3]Includes all SNPs included in activation and deactivation scores, allele associated with longer exposure to drug coded as 1, HR<1 expected

[4]Adjusted for age at diagnosis and tumor grade

Using logistic regression, completion of 6 or more cycles of cyclophosphamide was evaluated as a proxy for toxicity (Table 12). No association was observed between any of the gene scores and completion of 6 or more cycles of cyclophosphamide treatment. We would expect that those who activated the drug more quickly or deactivated the drug more slowly, resulting in longer exposure to the active metabolite, may be more likely to experience a toxicity event and may be more likely to complete less than 6 cycles of chemotherapy treatment with cyclophosphamide.

Table 12: Association Between Cyclophosphamide Genes Scores and Cycles of Cyclophosphamide Completed

Role in Metabolism	**OR (95% CI)[3]**	***P***
Activation[1]	0.96 (0.82, 1.12)	0.58
Deactivation[2]	0.97 (0.88, 1.06)	0.45
Total Score[1,2]	1.02 (0.94, 1.10)	0.68

[1]Includes SNPs from CYP genes
[2]Includes SNPs from ALDH and GST genes
[3]Adjusted for age at diagnosis and tumor grade

The association between cyclophosphamide gene scores and breast cancer survival among those who took cyclophosphamide was further evaluated through additional adjustment for the number of cycles of cyclophosphamide taken (Table 13). The observed associations between cyclophosphamide gene scores and survival outcomes were not materially changed after adjustment for number of cycles completed.

Table 13: Association Between Cyclophosphamide Gene Scores and Survival Outcomes Additionally Adjusted for Cycles of Cyclophosphamide Completed

Role in Metabolism	# of SNPs	N	Disease-Free Survival		Overall Survival	
			HR (95% CI)[4]	*P*	HR (95% CI)[4]	*P*
Activation[1]	4	889	0.98 (0.81, 1.20)	0.87	0.90 (0.73, 1.12)	0.34
Deactivation[2]	7	878	1.05 (0.94, 1.19)	0.38	1.01 (0.89, 1.15)	0.87
Total Score[3]	11	878	0.96 (0.87, 1.06)	0.42	0.97 (0.87, 1.08)	0.56

[1]Includes SNPs from CYP genes, allele associated with faster metabolism coded as 1, HR<1 expected
[2]Includes SNPs from ALDH and GST genes, allele associated with faster metabolism coded as 1, HR>1 expected
[3]Includes all SNPs included in activation and deactivation scores, allele associated with longer exposure to drug coded as 1, HR<1 expected
[4]Adjusted for age at diagnosis, tumor grade, and number of cyclophosphamide cycles completed

We stratified our results by timing of survival events, those that occurred in the first three years of follow-up and those that occurred after 3 years, among all women and restricted to those who took cyclophosphamide (Table 14). No clear differences were observed between any of the scores among those who had events in the first 3 years and those who had later events.

Table 14: Association Between Cyclophosphamide Gene Scores Stratified by Early vs. Late Events Among All Participants and Only Those Who Took Cyclophosphamide

		Event <3 years				Event ≥3 years			
		Disease-Free Survival		Overall Survival		Disease-Free Survival		Overall Survival	
Role in Metabolism	**# of SNPs**	**HR (95% CI)[4]**	***P***	**HR (95% CI)[4]**	***P***	**HR (95% CI)[4]**	***P***	**HR (95% CI)[4]**	***P***
All Participants (N=1124)									
Activation[1]	4	1.00 (0.81, 1.24)	0.97	1.03 (0.76, 1.40)	0.83	1.01 (0.76, 1.35)	0.94	0.88 (0.71, 1.10)	0.27
Deactivation[2]	7	1.09 (0.96, 1.24)	0.17	1.06 (0.89, 1.28)	0.51	1.03 (0.87, 1.22)	0.75	1.01 (0.88, 1.16)	0.92
Total Score[3]	11	0.94 (0.84, 1.05)	0.25	0.96 (0.82, 1.12)	0.63	0.98 (0.85, 1.13)	0.78	0.96 (0.86, 1.08)	0.53
Participants who Took Cyclophosphamide (N=884)									
Activation[1]	4	0.98 (0.77, 1.27)	0.90	0.96 (0.67, 1.38)	0.83	1.01 (0.73, 1.39)	0.96	0.86 (0.66, 1.13)	0.29
Deactivation[2]	7	1.09 (0.93, 1.27)	0.29	1.12 (0.90, 1.38)	0.32	1.01 (0.84, 1.22)	0.89	0.96 (0.82, 1.12)	0.61
Total Score[3]	11	0.94 (0.83, 1.07)	0.36	0.91 (0.76, 1.09)	0.33	0.99 (0.85, 1.16)	0.90	1.00 (0.87, 1.14)	0.98

[1]Includes SNPs from CYP genes, allele associated with faster metabolism coded as 1, HR<1 expected

[2]Includes SNPs from ALDH and GST genes, allele associated with faster metabolism coded as 1, HR>1 expected

[3]Includes all SNPs included in activation and deactivation scores, allele associated with longer exposure to drug coded as 1, HR<1 expected

[4]Adjusted for age at diagnosis and tumor grade

5-Fluorouracil

The association between each SNP in genes involved in metabolism of 5-fluorouracil and DFS and OS among all breast cancer participants was evaluated (Table 15). After adjustment for age at breast cancer diagnosis and tumor grade, no significant associations between any SNPs and DFS or OS were observed. Additional adjustment for education, BMI, menopausal status, tamoxifen use, ER status, PR status, HER2 status, radiotherapy, and mastectomy did not materially alter the observed associations.

When the results were stratified by whether 5-fluorouracil was taken, no significant associations were observed between SNPs in genes involved in the metabolism of 5-fluorouracil and DFS or OS among those who took 5-fluorouracil or those who did not (Table 16). No significant interactions were observed between those who took 5-fluorouracil and those who did not.

Table 15: Associations Between Individual SNPs in 5-Fluorouracil Metabolizing Genes and DFS and OS Among All Breast Cancer Cases

				All Participants							
				Disease-Free Survival				Overall Survival			
Gene	**rs ID**	**Effect Allele Frequency**	**N**	**HR (95% CI)[1]**	***P***	**HR (95% CI)[2]**	***P***	**HR (95% CI)[1]**	***P***	**HR (95% CI)[2]**	***P***
Activation[3]											
TYMP	rs11479	A=0.21	3736	0.99 (0.85, 1.16)	0.95	0.99 (0.85, 1.16)	0.90	1.00 (0.85, 1.18)	1.00	0.98 (0.83, 1.16)	0.84
UMPS	rs1801019	C=0.18	3736	0.92 (0.78, 1.08)	0.31	0.93 (0.79, 1.09)	0.36	0.95 (0.81, 1.13)	0.59	0.97 (0.82, 1.15)	0.71
UMPS	rs3772809	A=0.94	3739	1.11 (0.84, 1.47)	0.47	1.15 (0.87, 1.52)	0.33	1.17 (0.87, 1.57)	0.31	1.20 (0.89, 1.62)	0.22
Deactivation[4]											
DPYD	rs17376848	G=0.10	1124	1.00 (0.64, 1.56)	0.98	1.00 (0.63, 1.58)	1.00	1.24 (0.78, 1.96)	0.36	1.20 (0.75, 1.93)	0.45
DPYD	rs1801159	T=0.73	3736	0.99 (0.87, 1.14)	0.91	1.00 (0.87, 1.15)	0.98	0.97 (0.85, 1.12)	0.73	0.98 (0.85, 1.13)	0.81
DPYD	rs1801265	A=0.91	3736	1.02 (0.82, 1.27)	0.85	1.05 (0.84, 1.30)	0.69	1.06 (0.84, 1.34)	0.64	1.06 (0.84, 1.34)	0.63
DPYD	rs72728438	T=0.77	1124	1.24 (0.92, 1.67)	0.16	1.24 (0.92, 1.67)	0.16	1.27 (0.92, 1.76)	0.15	1.28 (0.92, 1.77)	0.15
Response[5]											
MTHFR	rs1801131	G=0.18	3739	0.91 (0.77, 1.08)	0.28	0.90 (0.76, 1.07)	0.24	0.91 (0.76, 1.08)	0.29	0.89 (0.75, 1.07)	0.22
MTHFR	rs1801133	G=0.56	1124	0.93 (0.71, 1.22)	0.61	0.93 (0.71, 1.23)	0.61	0.94 (0.71, 1.25)	0.66	0.93 (0.69, 1.24)	0.60
MTHFR	rs2274976	C=0.91	3736	1.00 (0.81, 1.24)	0.97	1.02 (0.82, 1.26)	0.87	0.97 (0.78, 1.21)	0.81	0.98 (0.78, 1.22)	0.85
TYMS	rs2847153	A=0.36	1145	1.18 (0.93, 1.50)	0.17	1.12 (0.88, 1.43)	0.36	0.95 (0.73, 1.23)	0.70	0.90 (0.69, 1.18)	0.46
TYMS	rs2853533	G=0.49	1124	0.86 (0.66, 1.12)	0.26	0.90 (0.69, 1.18)	0.44	0.95 (0.72, 1.25)	0.70	0.99 (0.75, 1.32)	0.95

[1]Adjusted for age at diagnosis and tumor grade

[2]Adjusted for age at diagnosis, tumor grade, education, BMI, menopausal status, tamoxifen use, ER status, PR status, HER2 status, radiotherapy, and mastectomy

[3]Allele associated with faster metabolism coded as 1, HR<1 expected

[4]Allele associated with faster metabolism coded as 1, HR>1 expected

[5]Allele associated with better survival coded as 1, HR<1 expected

Table 16: Associations Between Individual SNPs in 5-Fluorouracil Metabolizing Genes and DFS and OS Among All Breast Cancer Cases Stratified by Whether 5-Fluorouracil was Taken

		Participants who Took 5-Fluorouracil				Participants who Did Not Take 5-Fluorouracil					
		Disease-Free Survival		Overall Survival		Disease-Free Survival		Overall Survival		DFS	OS
Gene	**rs ID**	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***	P_{int}	P_{int}
Activation[3]											
TYMP	rs11479	1.04 (0.85, 1.27)	0.72	1.06 (0.86, 1.31)	0.59	0.92 (0.72, 1.17)	0.48	0.90 (0.69, 1.18)	0.45	0.47	0.37
UMPS	rs1801019	0.87 (0.71, 1.08)	0.22	0.88 (0.71, 1.10)	0.26	0.98 (0.76, 1.27)	0.88	1.08 (0.82, 1.41)	0.59	0.49	0.24
UMPS	rs3772809	1.07 (0.75, 1.52)	0.71	1.06 (0.73, 1.52)	0.77	1.18 (0.75, 1.85)	0.48	1.38 (0.83, 2.29)	0.22	0.70	0.37
Deactivation[4]											
DPYD	rs17376848	0.87 (0.51, 1.47)	0.60	1.27 (0.76, 2.13)	0.36	1.63 (0.66, 4.04)	0.29	1.36 (0.49, 3.75)	0.55	0.22	0.99
DPYD	rs1801159	1.07 (0.90, 1.28)	0.43	1.05 (0.87, 1.27)	0.58	0.90 (0.73, 1.12)	0.34	0.88 (0.71, 1.10)	0.28	0.22	0.22
DPYD	rs1801265	1.02 (0.77, 1.35)	0.90	1.03 (0.76, 1.40)	0.83	1.03 (0.72, 1.47)	0.86	1.11 (0.76, 1.61)	0.59	0.93	0.66
DPYD	rs72728438	1.20 (0.86, 1.67)	0.29	1.13 (0.79, 1.62)	0.51	1.33 (0.68, 2.60)	0.41	1.92 (0.86, 4.28)	0.11	0.71	0.19
Response[5]											
MTHFR	rs1801131	0.93 (0.75, 1.16)	0.51	0.88 (0.70, 1.11)	0.29	0.87 (0.67, 1.14)	0.31	0.95 (0.72, 1.24)	0.68	0.72	0.72
MTHFR	rs1801133	0.99 (0.73, 1.34)	0.96	1.00 (0.73, 1.37)	0.99	0.74 (0.39, 1.39)	0.35	0.73 (0.37, 1.43)	0.35	0.37	0.39
MTHFR	rs2274976	0.98 (0.74, 1.29)	0.87	0.93 (0.70, 1.24)	0.63	1.08 (0.77, 1.52)	0.66	1.04 (0.73, 1.48)	0.83	0.63	0.56
TYMS	rs2847153	1.11 (0.85, 1.46)	0.45	0.87 (0.64, 1.18)	0.37	1.45 (0.88, 2.41)	0.15	1.35 (0.79, 2.32)	0.27	0.36	0.15
TYMS	rs2853533	0.87 (0.64, 1.17)	0.34	1.02 (0.74, 1.39)	0.91	0.80 (0.45, 1.43)	0.46	0.68 (0.37, 1.26)	0.22	0.96	0.30

[1]Adjusted for age at diagnosis and tumor grade

[2]Adjusted for age at diagnosis, tumor grade, education, BMI, menopausal status, tamoxifen use, ER status, PR status, HER2 status, radiotherapy, and mastectomy

[3]Allele associated with faster metabolism coded as 1, HR<1 expected

[4]Allele associated with faster metabolism coded as 1, HR>1 expected

[5]Allele associated with better survival coded as 1, HR<1 expected

Each of the five gene scores created for SNPs in genes involved in 5-fluorouracil metabolism were evaluated for their association with DFS and OS among all breast cancer cases (Table 17). No significant associations were observed between the gene scores and survival outcomes among all breast cancer patients, although the total scores were in the expected direction (longer exposure to drug associated with better survival).

When stratified by whether 5-fluorouracil was taken, no significant associations were observed among the five gene scores and DFS among those who underwent treatment with 5-fluoruracil (Table 17). The total genes scores were more strongly associated with better OS among those who took 5-fluorouracil compared to those who did not, although the association was not statistically significant. Additionally, no significant interactions were observed between those who took 5-fluorouracil and those who did not.

Table 17: Association between 5-Fluorouracil Gene Scores and Breast Cancer Outcomes Among All Breast Cancer Cases and Stratified by Whether 5-Fluorouracil was Taken

Role in Metabolism	# of SNPs	N	Disease-Free Survival HR (95% CI)[6]	*P*	Overall Survival HR (95% CI)[6]	*P*
All Participants						
Activation[1]	3	3736	0.98 (0.89, 1.08)	0.69	1.00 (0.90, 1.11)	0.96
Deactivation[2]	4	1124	1.00 (0.84, 1.19)	0.98	1.01 (0.84, 1.21)	0.95
Other[3]	5	1124	0.97 (0.82, 1.15)	0.71	0.91 (0.76, 1.09)	0.29
Total Score[4]	7	1124	0.99 (0.86, 1.13)	0.87	0.91 (0.78, 1.05)	0.19
Total Score[5]	12	1124	0.98 (0.88, 1.09)	0.70	0.90 (0.80, 1.01)	0.08
Participants Who Took 5-Fluorouracil						
Activation[1]	3	2625	0.97 (0.85, 1.11)	0.68	0.98 (0.86, 1.12)	0.79
Deactivation[2]	4	882	1.01 (0.84, 1.23)	0.88	1.03 (0.84, 1.27)	0.75
Other[3]	5	882	0.97 (0.80, 1.17)	0.75	0.93 (0.76, 1.13)	0.45
Total Score[4]	7	882	0.98 (0.84, 1.13)	0.74	0.87 (0.74, 1.02)	0.09
Total Score[5]	12	882	0.97 (0.86, 1.10)	0.64	0.88 (0.77, 1.00)	0.06
Participants who Did Not Take 5-Fluorouracil						
Activation[1]	3	1111	0.98 (0.83, 1.14)	0.76	1.03 (0.87, 1.22)	0.76
Deactivation[2]	4	242	0.96 (0.65, 1.40)	0.82	0.92 (0.62, 1.37)	0.68
Other[3]	5	242	0.98 (0.66, 1.44)	0.90	0.85 (0.56, 1.29)	0.45
Total Score[4]	7	242	1.11 (0.79, 1.56)	0.56	1.05 (0.73, 1.50)	0.79
Total Score[5]	12	242	1.05 (0.81, 1.35)	0.72	0.96 (0.73, 1.26)	0.77

[1]Includes SNPs from TYMP and UMPS genes, allele associated with faster metabolism coded as 1, HR<1 expected
[2]Includes SNPs from DPYD gene, allele associated with faster metabolism coded as 1, HR>1 expected
[3]Includes SNPs from TYMS and MTHFR genes, allele associated with better survival coded as 1, HR<1 expected
[4]Includes all SNPs included in activation and deactivation scores, allele associated with longer exposure to drug coded as 1, HR<1 expected
[5]Includes all SNPs included in activation, deactivation, and other scores, allele associated with longer exposure to drug coded as 1, HR<1 expected
[6]Adjusted for age at diagnosis and tumor grade

Using logistic regression, completion of 6 or more cycles of 5-fluorouracil was evaluated as a surrogate for toxicity (Table 18). We would expect that those who activated the drug more quickly or deactivated the drug more slowly, resulting in longer exposure to the active metabolite, may be more likely to experience a toxicity event and may be more likely to complete less than 6 cycles of chemotherapy treatment with 5-fluorouracil. Among breast cancer participants who underwent treatment with 5-fluorouracil, there was a 10% decrease in likelihood of completing 6 or more cycles of 5-fluorouracil for each additional SNP associated with increased activation of 5-fluorouracil (OR=0.90, 95% CI: 0.81, 1.00, p=0.04). The 5-fluorouracil response gene score was not associated with number of cycles completed and neither was the combined gene score which included these SNPs. No other associations between 5-fluorouracil gene scores and number of cycles of 5-fluorouracil completed were observed.

Table 18: Association Between 5-Fluorouracil Genes Scores and Cycles of 5-Fluorouracil Completed

Role in Metabolism	**OR (95% CI)[4]**	***P***
Activation[1]	0.90 (0.81, 1.00)	0.04
Deactivation[2]	1.05 (0.90, 1.23)	0.51
Response[3]	1.08 (0.93, 1.26)	0.31
Total Score[1,2]	0.97 (0.87, 1.09)	0.62
Total Score[1,2,3]	1.01 (0.92, 1.11)	0.83

[1]Includes SNPs from TYMP and UMPS genes
[2]Includes SNPs from DPYD gene
[3]Includes SNPs from TYMS and MTHFR genes
[4]Adjusted for age at diagnosis and tumor grade

The association between 5-fluorouracil gene scores and breast cancer survival among those who took 5-fluorouracil was further evaluated through additional adjustment for the number of cycles of 5-fluorouracil taken (Table 19). The observed associations between 5-fluorouracil gene scores and survival outcomes were not materially changed after adjustment for number of cycles completed.

Table 19: Association Between 5-Fluorouracil Gene Scores and Survival Outcomes Additionally Adjusted for Cycles of 5-Fluorouracil Completed

			Disease-Free Survival		**Overall Survival**	
Role in Metabolism	**# of SNPs**	**N**	**HR (95% CI)[6]**	***P***	**HR (95% CI)[6]**	***P***
Activation[1]	3	2623	0.97 (0.85, 1.10)	0.60	0.99 (0.86, 1.13)	0.83
Deactivation[2]	4	881	1.02 (0.84, 1.23)	0.86	1.03 (0.83, 1.27)	0.79
Response[3]	5	881	0.97 (0.80, 1.18)	0.79	0.93 (0.76, 1.14)	0.51
Total Score[4]	7	881	0.98 (0.84, 1.13)	0.74	0.88 (0.75, 1.04)	0.13
Total Score[5]	12	881	0.97 (0.86, 1.10)	0.65	0.89 (0.78, 1.02)	0.09

[1]Includes SNPs from TYMP and UMPS genes, allele associated with faster metabolism coded as 1, HR<1 expected
[2]Includes SNPs from DPYD gene, allele associated with faster metabolism coded as 1, HR>1 expected
[3]Includes SNPs from TYMS and MTHFR genes, allele associated with better survival coded as 1, HR<1 expected
[4]Includes all SNPs included in activation and deactivation scores, allele associated with longer exposure to drug coded as 1, HR<1 expected
[5]Includes all SNPs included in activation, deactivation, and other scores, allele associated with longer exposure to drug coded as 1, HR<1 expected
[6]Adjusted for age at diagnosis and tumor grade

We stratified our results by timing of survival events, those that occurred in the first three years of follow-up and those that occurred after 3 years, among all women and restricted to those

who took 5-fluorouracil (Table 20). The 5-fluorouracil gene scores were both associated with better OS for events which occurred after the first 3 years, particularly among those who took 5-fluorouracil (total score including SNPs in activating and deactivating genes only HR=0.76, 95% CI: 0.62, 0.93, p=0.008, total score including all SNPs HR=0.78, 95% CI: 0.67, 0.92, p=0.004) and the association was significant even after correction for multiple comparisons. No association was observed in those who had events in the first 3 years following cancer diagnosis.

Table 20: Association Between 5-fluorouracil Gene Scores Stratified by Early vs. Late Events Among All Participants and Only Those who Took 5-Fluorouracil

		Event <3 years				Event ≥3 years			
		Disease-Free Survival		Overall Survival		Disease-Free Survival		Overall Survival	
Role in Metabolism	**# of SNPs**	**HR (95% CI)[6]**	***P***	**HR (95% CI)[6]**	***P***	**HR (95% CI)[6]**	***P***	**HR (95% CI)[6]**	***P***
All Participants									
Activation[1]	3	0.98 (0.87, 1.11)	0.80	0.98 (0.84, 1.15)	0.81	0.96 (0.81, 1.15)	0.68	1.02 (0.89, 1.18)	0.78
Deactivation[2]	4	0.99 (0.80, 1.23)	0.93	0.90 (0.66, 1.21)	0.48	1.01 (0.76, 1.36)	0.92	1.07 (0.85, 1.35)	0.56
Response[3]	5	1.06 (0.86, 1.31)	0.57	0.94 (0.70, 1.26)	0.68	0.81 (0.60, 1.08)	0.15	0.89 (0.71, 1.12)	0.32
Total Score[4]	7	1.00 (0.85, 1.19)	0.96	1.05 (0.83, 1.33)	0.71	0.97 (0.77, 1.22)	0.78	0.83 (0.69, 1.00)	0.05
Total Score[5]	12	1.03 (0.90, 1.18)	0.68	1.00 (0.83, 1.22)	0.97	0.90 (0.74, 1.08)	0.25	0.84 (0.73, 0.98)	0.02
Participants who Took 5-Fluorouracil									
Activation[1]	3	0.98 (0.83, 1.15)	0.76	1.01 (0.82, 1.25)	0.92	0.96 (0.78, 1.19)	0.73	0.96 (0.80, 1.14)	0.62
Deactivation[2]	4	1.03 (0.80, 1.32)	0.84	0.87 (0.61, 1.23)	0.42	0.99 (0.73, 1.34)	0.94	1.14 (0.88, 1.48)	0.32
Response[3]	5	1.11 (0.87, 1.42)	0.40	1.05 (0.75, 1.47)	0.79	0.77 (0.57, 1.06)	0.11	0.87 (0.68, 1.12)	0.27
Total Score[4]	7	0.96 (0.79, 1.16)	0.66	1.10 (0.84, 1.43)	0.50	1.01 (0.80, 1.28)	0.92	0.76 (0.62, 0.93)	0.008
Total Score[5]	12	1.02 (0.87, 1.19)	0.85	1.09 (0.87, 1.35)	0.47	0.91 (0.75, 1.11)	0.36	0.78 (0.67, 0.92)	0.004

[1]Includes SNPs from TYMP and UMPS genes, allele associated with faster metabolism coded as 1, HR<1 expected

[2]Includes SNPs from DPYD gene, allele associated with faster metabolism coded as 1, HR>1 expected

[3]Includes SNPs from TYMS and MTHFR genes, allele associated with better survival coded as 1, HR<1 expected

[4]Includes all SNPs included in activation and deactivation scores, allele associated with longer exposure to drug coded as 1, HR<1 expected

[5]Includes all SNPs included in activation, deactivation, and other scores, allele associated with longer exposure to drug coded as 1, HR<1 expected

[6]Adjusted for age at diagnosis and tumor grade

Joint Effect of Cyclophosphamide Gene Score and 5-Fluorouracil Gene Score

Approximately 65% of our population underwent chemotherapy with both cyclophosphamide and 5-fluorouracil. We further evaluated the joint effect of the two total scores which we created (Table 21). For comparability, we used the 5-fluorouracil gene score which included SNPs in genes involved in activation and deactivation only. No joint effect of the two gene scores was observed. No difference in survival was observed between all participants and those participants who took both drugs.

Table 21: Joint Effect of the Cyclophosphamide Gene Score and the 5-Fluorouracil Gene Score

	Disease-Free Survival			**Overall Survival**		
	Cyclophosphamide Score (HR[1] (95%CI))			**Cyclophosphamide Score (HR[1] (95%CI))**		
5-Fluorouracil Score	4.00-10.59	10.60-12.33	12.34-18.50	4.00-10.59	10.60-12.33	12.34-18.50
Overall (n=1124)						
0.0-4.0	1.00 (reference)	1.52 (0.74, 3.13)	0.98 (0.45, 2.12)	1.00 (reference)	1.00 (0.45, 2.24)	1.11 (0.52, 2.40)
4.0-5.0	2.19 (1.13, 4.26)	1.43 (0.72, 2.82)	1.09 (0.52, 2.30)	1.90 (0.95, 3.80)	1.19 (0.58, 2.45)	1.19 (0.55, 2.58)
5.0-11.0	0.91 (0.39, 2.12)	0.98 (0.45, 2.13)	1.07 (0.52, 2.23)	0.95 (0.41, 2.16)	0.60 (0.25, 1.44)	0.90 (0.42, 1.95)
Those who Took Cyclophosphamide and 5-Fluorouracil (n=845)						
0.0-4.0	1.00 (reference)	1.42 (0.62, 3.23)	0.94 (0.39, 2.24)	1.00 (reference)	1.08 (0.43, 2.68)	1.17 (0.50, 2.77)
4.0-5.0	1.97 (0.93, 4.19)	1.57 (0.75, 3.30)	1.08 (0.48, 2.47)	1.57 (0.69, 3.56)	1.37 (0.62, 3.02)	1.12 (0.47, 2.64)
5.0-11.0	0.68 (0.24, 1.93)	0.59 (0.22, 1.58)	1.05 (0.46, 2.40)	0.93 (0.36, 2.41)	0.29 (0.08, 1.05)	0.91 (0.39, 2.15)

[1]Adjusted for age at diagnosis and tumor grade

Conclusions/Discussion

In this study, we evaluated 11 SNPs in genes involved in cyclophosphamide metabolism and 12 SNPs in genes involved in 5-fluorouracil metabolism and response. We found that the C allele in the rs2228100 SNP in the ALDH3A1 gene and the G allele in the rs1695 SNP in the GSTP1 gene were associated with better DFS, though associations were not significant after adjustment for multiple comparisons. The A allele in the rs4986893 SNP in the CYP2C19 gene was associated with poorer DFS and OS, though the association was not significant after adjustment for multiple comparisons. No significant interactions were observed between those who took cyclophosphamide and those that did not. The cyclophosphamide gene score was not associated with DFS or OS. No individual alleles in genes involved in 5-fluorouracil metabolism were associated with DFS or OS. The 5-fluorouracil gene score was associated with improved OS, particularly among those who took 5-fluorouracil who survived at least three years without an event.

The effects of germline genetic variation in enzymes involved in chemotherapy metabolism have the potential to explain differences in response to chemotherapy drugs and cancer prognosis. However, the effect of a single polymorphism may not be great enough to observe an association. Our pathway approach increased power to detect associations between SNPs in these genes and breast cancer survival outcomes. We chose *a priori* to incorporate all potentially functioning SNPs investigated in this study by creating gene score variables. While previous studies have shown an association between individual SNPs and chemotherapy response or breast cancer outcomes, the evidence for most of the SNPs included in this study was inconsistent. Inclusion of SNPs in the gene score which do not affect chemotherapy response and

breast cancer outcomes would decrease our power to detect an association between our gene score and survival. Furthermore, SNPs in the promoter regions of our genes of interest were not included in this study. Future studies should also consider inclusion of these SNPs.

We did not observe significant differences in the cyclophosphamide gene scores we created between those who took cyclophosphamide and those who did not. The genes which we evaluated in this study are also involved in other biological processes, such as cell proliferation and angiogenic activity, some of which may have an effect of tumor progression and survival. Therefore, our scores may not only be measuring potential effects on chemotherapy activation and degradation but also other biological function of the genes. Additionally, there are other factors that could affect drug metabolism and potentially gene expression levels including comorbidities, medications, and dietary factors. While some of these factors were formally tested as potential confounders, we were unable to account for all of these factors in our analyses.

The 5-fluorouracil total gene scores were marginally associated with better overall survival, particularly among those participants who were treated with 5-fluorouracil. As expected, those with a longer exposure period to the active metabolite had better outcomes.

We included all breast cancer types in order to maximize sample size. We further adjusted our results for ER, PR, and HER2 status, as well as Tamoxifen use, which is used to treat hormone receptor positive cancers and is associated with breast cancer survival, and there was no difference in the observed associations. Further studies which could evaluate these associations by molecular intrinsic subtypes, particularly in TNBC patients and patients diagnosed with late stage breast cancer where chemotherapy is standard, are needed.

CHAPTER V

TUMOR-LEVEL EXPRESSION OF CHEMOTHERAPY METABOLIZING GENES AND TRIPLE-NEGATIVE BREAST CANCER OUTCOMES

Aim 3-Specific Methods

As previously mentioned, gene expression data was $\log_2$ transformed prior to analysis to account for non-normal data distributions. Gene expression data was analyzed as a continuous measure and as a categorical variable with a median cut point and a tertile cut point.

The potential effect of gene expression levels of genes involved in cyclophosphamide metabolism was hypothesized based on knowledge of cyclophosphamide metabolism as well as results from previous studies.[24] As previously discussed, cyclophosphamide is activated in the liver by cytochrome P450 genes prior to entering the cell; therefore, we expect that the breast tumor level expression of these genes will have little or no effect on cyclophosphamide metabolism and breast cancer survival outcomes. The aldehyde dehydrogenase genes and glutathione-S-transferase genes are involved in clearance of the active cyclophosphamide metabolite; therefore, we expect high expression of these genes to be associated with shorter exposure to the active drug and worse survival compared to those with low expression.

The potential effect of gene expression levels for genes involved in 5-fluorouracil metabolism was hypothesized based on knowledge of 5-fluorouracil metabolism as well as results from previous studies.[24] As previously discussed, TYMP and UMPS are involved in activation of 5-fluorouracil to the active metabolite in the cell; therefore we expect higher expression of these genes to be associated with longer exposure to the active drug and better

survival compared to those with low expression. DPYD clears (deactivates) the active form of 5-fluorouracil; therefore we expect those with high DPYD expression to have shorter exposure to treatment and worse survival. TYMS is the target of 5-fluorouracil and 5-fluorouracil works through inhibiting this gene; therefore, we expect that those with high TYMS expression would have a decreased response and worse survival compared with those with low expression. MTHFR catalyzes the conversion of methylenetetrahydrofolate to methyltetrahydrofolate, the latter acts as a cofactor in the inhibition of TYMS; therefore, we would expect that those with low MTHFR expression would have increased response to treatment and better survival compared with those with low expression.

Genes which were expressed in more than 85% of tumor samples were analyzed as continuous variables, as well as categorical variables using median and tertile cut points. Genes expressed in less than 50% of tumor samples were only analyzed as expressed versus not expressed.

Five genes were included in the cyclophosphamide gene score, all of which are involved in cyclophosphamide deactivation; those with expression levels above the median were coded as 1 and those with expression levels below the median were coded as 0. For genes with expression in less than 50% of the tumor samples, those with expression were coded as 1 and those with no expression were coded as 0. This resulted in a score with a possible range of 0 to 5. Those with higher scores clear the active metabolite of cyclophosphamide faster than those with lower scores. The top and bottom ends of the score were collapsed due to low sample size, resulting in a score with a range of 1 to 3 (Table 22).

Table 22: Creation of Cyclophosphamide Gene Expression Score Based on Median Cut Points

Original Score	Collapsed Score	Frequency	Percent
0	**1**	25	6.0
1	**1**	130	31.1
2	**2**	150	35.9
3	**3**	83	19.9
4	**3**	28	6.7
5	**3**	2	0.5

Five genes were included in the 5-fluorouracil gene score, all of which were expressed in >85% of tumor samples. Two of the genes are involved in activation of 5-fluorouracil to the active metabolite, one is involved in clearance of the active metabolite, and two are involved in response. For the activating genes, those with expression levels above the median were coded as 0 and those with expression levels below the median were coded as 1. For the deactivating genes, those with expression levels above the median were coded as 1 and those with expression levels below the median were coded as 0. The response genes were coded so that those with worse survival expected (based on expression level) were coded as 1 and those who were expected to respond better as 0. This resulted in a score with a possible range of 0 to 5 where those with higher scores are faster deactivators, slower activators, and worse responders. The top and bottom ends of the score were collapsed due to low sample size, resulting in a score with a range of 1 to 4 (Table 23).

Table 23: Creation of 5-Fluorouracil Gene Expression Score Based on Median Cut Points

Original Score	Collapsed Score	Frequency	Percent
0	**1**	7	1.7
1	**1**	60	14.4
2	**2**	142	34.0
3	**3**	138	33.0
4	**4**	63	15.1
5	**4**	8	1.9

Results

Among TNBC participants in the SBCSS cohort, 67 recurrences/breast cancer deaths and 76 deaths were documented over a median follow-up of 5.3 years (range: 0.7-8.9 years). Advanced stage disease and radiotherapy treatment were inversely associated with 5-year DFS and OS (Table 24). Chemotherapy was inversely associated with DFS rate. Grade was inversely associated with OS rate.

Table 24: Demographic and Clinical Predictors for Breast Cancer Survival for TNBC Cases in the SBCSS

		Disease-Free Survival			Overall Survival		
Characteristics	**N**	**Events, No.**	**5-Yr Survival Rate, %[1]**	***P***	**Deaths, No.**	**5-Yr Survival Rate, %[1]**	***P***
Age at diagnosis, y							
<40	29	4	84.3		5	82.8	
40-49	154	23	84.6	0.85	24	87.5	0.18
50-59	108	16	84.5		15	87.9	
≥60	127	24	79.6		32	80.3	
Education							
Elementary School or Less	60	13	75.7		17	78.1	
Middle School	140	27	80.3	0.18	25	85.7	0.14
High or Vocational School	151	16	89.3		20	88.5	
College or University	67	11	81.2		14	81.9	
Income (yuan/person/month)							
<500	53	12	75.2		14	81.1	
500 - <700	60	15	74.9		15	81.5	
700 - <1000	127	20	83.1	0.06	23	83.3	0.32
1000 - <2000	133	17	86.8		18	88.6	
≥2000	44	3	92.1		6	88.4	
Body Mass Index (kg/m^2)							
<25	271	43	83.2		46	86.7	
25-29.99	121	18	84.1	0.70	24	82.5	0.68
≥30	26	6	75.4		6	80.3	
Menopausal Status							
Premenopausal	193	27	85.7	0.37	28	87.9	0.08
Postmenopausal	225	40	80.8		48	82.6	
TNM Stage							
0-I	137	12	90.3		14	90.4	
IIA	145	18	87.3		20	89.6	
IIB	85	20	73.5	<.0001	25	75.1	<.0001
III-IV	38	16	58.2		16	68.4	
Unknown	13	1	91.7		1	92.3	
Grade							
1	50	5	89.0		3	94.0	
2	132	19	83.9	0.28	23	87.1	0.04
3	236	43	81.2		50	82.0	
Chemotherapy							
Yes	390	59	84.1	0.04	67	86.3	0.06
No	28	8	68.1		9	67.9	
Radiotherapy							
Yes	103	27	71.8	0.0009	28	75.5	0.002
No	315	40	86.7		48	88.1	
Mastectomy							
Yes	399	63	83.3	0.36	73	85.1	0.73
No	19	4	76.1		3	84.2	
No. of Live Births							
0	2	0	100.0		0	100	
1	268	39	84.7	0.54	40	87.5	0.07
2	79	16	78.8		18	81.0	
≥3	53	11	76.9		16	77.4	
Family History of BC							
Yes	29	6	76.2	0.37	5	85.9	0.93
No	389	61	83.5		71	84.9	

[1]Survival rate calculated using life table analysis method

The gene expression levels of each of the cyclophosphamide and 5-fluorouracil metabolizing genes of interest in our study population are shown in Table 25 and Table 26, respectively.

Table 25: Expression Levels of Cyclophosphamide Metabolizing Genes Among TNBC Participants in the SBCSS Cohort

Gene	% Expressed	Mean (SD)	Median	Minimum	Maximum
CYP2B6	9.3%	3.9 (1.8)	4.0	1.0	7.4
CYP3A4	0.0%	n/a	n/a	n/a	n/a
CYP2C9	data not available				
CYP3A5	8.6%	2.8 (1.5)	2.7	1.0	5.7
CYP2A6	9.6%	6.5 (3.3)	6.2	1.0	14.7
CYP2C8	12.0%	2.5 (1.0)	2.3	1.0	4.5
CYP2C19	0.2%	1.0 (n/a)	1.0	1.0	1.0
ALDH1A1	95.5%	7.1 (1.5)	7.2	1.0	11.8
ALDH3A1	data not available				
GSTP1	100.0%	10.9 (1.0)	11.1	6.6	13.6
GSTA1	10.0%	3.6 (1.7)	3.2	1.0	6.9
GSTT1	44.0%	6.4 (1.4)	6.7	1.0	8.9
GSTM1	37.6%	6.9 (2.3)	6.8	1.0	13.8

Statistics among those with expression only

Table 26: Expression Levels of 5-Fluorouracil Metabolizing Genes Among TNBC Participants in the SBCSS Cohort

Gene	% Expressed	Mean (SD)	Median	Minimum	Maximum
DPYD	96.7%	7.8 (0.9)	7.9	2.0	10.1
TYMS	96.7%	8.8 (1.3)	8.8	3.3	12.1
MTHFR	95.5%	6.8 (0.9)	6.9	2.3	9.2
TYMP	100.0%	10.7 (1.0)	10.7	5.6	14.4
UMPS	90.2%	6.2 (0.9)	6.2	2.0	8.4

Statistics among those with expression only

ALDH1A1, GSTP1, DPYD, TYMS, MTHFR, TYMP, and UMPS were expressed in the large majority (≥90%) of TNBC tumor tissue samples in our study. GSTA1, GSTM1, and GSTT1 were expressed in fewer than 50% of samples. As expected, the CYP genes were not expressed or expressed at very low levels in the tumor tissue. Therefore, the tumor level expression of the CYP genes will not be investigated in this study.

Cyclophosphamide

Participants with expression levels of ALDH1A1 below the median were more likely to have a higher grade tumor and more likely to have markers of the basal-like subtype compared to those with expression levels above the median (Table 27). Participants with expression levels of GSTP1 above the median were more likely to have a higher grade tumor and more likely to have markers of the basal-like subtype compared to those with expression levels below the median. Those with expression levels of GSTP1 above the median were also less likely to have received chemotherapy, particularly cyclophosphamide. Expression of GSTA1 was associated with higher tumor grade. Expression of GSTT1 was associated with higher tumor grade. Lack of expression of GSTA1 or GSTM1 was associated with markers of the basal-like subtype.

Dual specificity phosphatase 4 (DUSP4) expression shows the strongest association with survival in this study population; this association was previously reported.[169] We examined the correlation between the cyclophosphamide metabolizing genes being investigated in this study and DUSP4 expression (Table 28). ALDH1A1 and GSTT1 expressions were positively correlated with DUSP4 expression ($p<.0001$ and $p=0.049$, respectively) and GSTP1 and GSTA1 were negatively associated with DUSP4 expression ($p=0.0003$ and $p=0.002$), respectively. No correlation between DUSP4 expression and GSTM1 expression was observed.

Table 27: Clinical and Treatment Factors by Expression of Cyclophosphamide Metabolizing Genes Among TNBC Participants in the SBCSS Cohort

	ALDH1A1			GSTP1			GSTA1			GSTM1			GSTT1		
	<Median N(%[1])	≥Median N(%[1])	*P*	<Median N(%[1])	≥Median N(%[1])	*P*	Not Expressed N(%[1])	Expressed N(%[1])	*P*	Not Expressed N(%[1])	Expressed N(%[1])	*P*	Not Expressed N(%[1])	Expressed N(%[1])	*P*
N	209	209		209	209		376	42		261	157		234	184	
TNM Stage			0.65			0.46			0.16			0.17			0.75
0-I	67 (48.9)	70 (21.1)		70 (51.1)	67 (48.9)		123 (89.8)	14 (10.2)		76 (55.5)	61 (44.5)		77 (56.2)	60 (43.8)	
IIA	70 (48.3)	75 (51.7)		75 (51.7)	70 (48.3)		133 (91.7)	12 (8.3)		98 (67.6)	47 (32.4)		80 (55.2)	65 (44.8)	
IIB	48 (56.5)	37 (43.5)		37 (43.5)	48 (56.5)		71 (83.5)	14 (16.5)		56 (65.9)	29 (34.1)		49 (57.7)	36 (42.4)	
III-IV	19 (50.0)	19 (50.0)		22 (57.9)	16 (42.1)		36 (94.7)	2 (5.3)		23 (60.5)	15 (39.5)		18 (47.4)	20 (52.6)	
Missing	5	8		5	8		13	0		8	5		10	3	
Grade			<.0001			<.0001			0.02			0.85			0.04
1	12 (24.0)	38 (76.0)		30 (60.0)	20 (40.0)		49 (98.0)	1 (2.0)		30 (60.0)	20 (40.0)		35 (70.0)	15 (30.0)	
2	56 (42.4)	76 (57.6)		84 (63.6)	48 (36.4)		123 (93.2)	9 (6.8)		81 (61.4)	51 (38.6)		65 (49.2)	67 (50.8)	
3	141 (59.8)	95 (40.3)		95 (40.3)	141 (59.8)		204 (86.4)	32 (13.6)		150 (63.6)	86 (36.4)		134 (56.8)	102 (43.2)	
Chemotherapy	198 (50.8)	192 (49.2)	0.24	200 (51.3)	190 (48.7)	0.05	351 (90.0)	39 (10.0)	0.90	244 (62.6)	146 (37.4)	0.85	216 (55.4)	174 (44.6)	0.36
Cyclophosphamide	151 (51.0)	145 (49.0)	0.52	158 (53.4)	138 (46.6)	0.03	268 (90.5)	28 (9.5)	0.53	186 (62.8)	110 (37.2)	0.79	163 (55.1)	133 (44.9)	0.56
Radiotherapy	59 (57.3)	44 (42.7)	0.09	50 (48.5)	53 (51.5)	0.73	89 (86.4)	14 (13.6)	0.17	65 (63.1)	38 (36.9)	0.87	58 (56.3)	45 (43.7)	0.94
Mastectomy	201 (50.4)	198 (49.6)	0.48	198 (49.6)	201 (50.4)	0.48	358 (89.7)	41 (10.3)	0.48	250 (62.7)	149 (37.3)	0.68	223 (55.9)	176 (44.1)	0.86
Subtype Classification			<.0001			<.0001			<.0001			0.01			0.38
Basal-like	124 (71.3)	50 (28.7)		56 (32.2)	118 (67.8)		141 (81.0)	33 (19.0)		108 (62.1)	66 (37.9)		98 (56.3)	76 (43.7)	
Her-2 Enriched	21 (35.0)	39 (65.0)		30 (50.0)	30 (50.0)		57 (95.0)	3 (5.0)		38 (63.3)	22 (36.7)		27 (45.0)	33 (55.0)	
Luminal A	32 (29.1)	78 (70.9)		75 (68.2)	35 (31.8)		110 (100.0)	0 (0.0)		76 (69.1)	34 (30.9)		65 (59.1)	45 (40.9)	
Luminal B	22 (53.7)	19 (46.3)		31 (75.6)	10 (24.4)		39 (95.1)	2 (4.9)		16 (39.0)	25 (61.0)		23 (56.1)	18 (43.9)	
Normal	10 (30.3)	23 (69.7)		17 (51.5)	16 (48.5)		29 (87.9)	4 (12.1)		23 (69.7)	10 (30.3)		21 (63.6)	12 (36.4)	

[1]Percents shown are row percentages

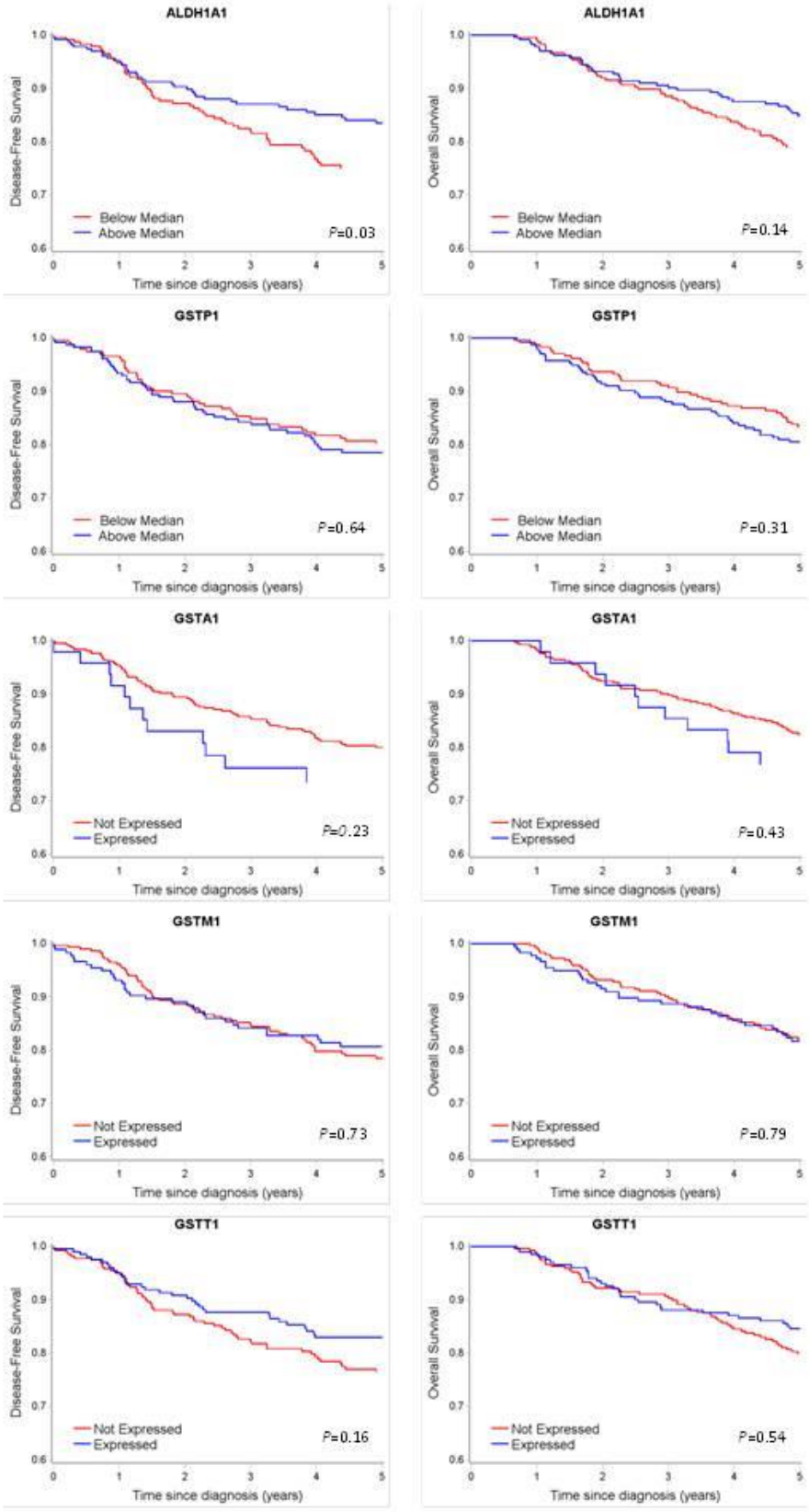
Median Values: ALDH1A1 - 7.140; GSTP1 - 11.120

Table 28: Correlation Between Expression of Cyclophosphamide Metabolizing Genes and DUSP4 Expression Among TNBC Participants in the SBCSS Cohort

	Correlation with DUSP4	
	r	***p***
ALDH1A1	0.28	<.0001
GSTP1	-0.18	0.0003
GSTA1	-0.15	0.002
GSTM1	0.01	0.81
GSTT1	0.10	0.049

Univariate analyses between survival, both disease-free and overall, and expression level of genes involved in cyclophosphamide among all TNBC patients are shown in Figure 2. ALDH1A1 expression above the median was significantly associated with better DFS (p=0.03) compared with expression levels below the median, but no significant difference was seen in OS (p=0.14). No difference was observed in DFS or OS by expression of GSTP1, GSTA1, GSTM1, or GSTT1.

Figure 2: Kaplan-Meier Curves For Disease-Free Survival and Overall Survival for Genes Involved in Cyclophosphamide Metabolism Among TNBC Participants in the SBCSS Cohort



As previously shown in our published paper,[121] among all TNBC patients, high ALDH1A1 expression, analyzed continuously, was significantly associated with DFS and OS, after adjustment for age at diagnosis (DFS HR=0.91, 95% CI: 0.82, 1.00, OS HR=0.89, 95% CI: 0.81, 0.98) (Table 29). This association was attenuated slightly after adjustment for tumor grade and markers of basal-like subtype. After further adjustment for DUSP4 expression, the association was further attenuated (DFS HR=0.97, 95% CI: 0.86, 1.08, OS HR=0.95, 95% CI: 0.86, 1.06). ALDH1A1 expression above the median was significantly associated with DFS, but not OS, compared to the lower median, after adjustment for age at diagnosis and tumor grade (DFS HR=0.60, 95% CI: 0.36, 1.00, OS HR=0.74, 95% CI: 0.46, 1.18). Further adjustment for markers of the basal-like subtype and DUSP4 expression attenuated the association with DFS (HR=0.70, 95% CI: 0.40, 1.21). However, when the highest tertile of ALDH1A1 gene expression was compared to the lowest tertile, the significant association with better DFS remained, even after adjustment for age at diagnosis, tumor grade, markers of the basal-like subtype, and DUSP4 expression (HR=0.41, 95% CI: 0.19, 0.88).

No association was observed between GSTP1, GSTA1, GSTM1, or GSTT1 expression and DFS or OS after adjustment for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression.

Next, the results were stratified by whether the drug of interest, cyclophosphamide, was taken (Table 30). No significant differences were seen in the associations between gene expression and survival outcomes based on whether cyclophosphamide was taken. However, among those who took cyclophosphamide, GSTP1 expression was associated with significantly better DFS (continuous HR=0.65, 95% CI: 0.47, 0.91).

Table 29: Disease-Free and Overall Survival by Expression of Genes Involved in Cyclophosphamide Metabolism Among TNBC Participants in the SBCSS Cohort

	5-yr Rate, %[1]	HR (95% CI)[2]	*P*	HR (95% CI)[3]	*P*	HR (95% CI)[4]	*P*	HR (95% CI)[5]	*P*
Disease-Free Survival									
ALDH1A1									
Continuous	83.0	0.91 (0.82, 1.00)	0.05	0.92 (0.83, 1.02)	0.10	0.93 (0.83, 1.03)	0.18	0.97 (0.86, 1.08)	0.55
<median	79.1	Reference		Reference		Reference		Reference	
≥median	86.7	0.55 (0.34, 0.91)	0.02	0.60 (0.36, 1.00)	0.05	0.65 (0.38, 1.11)	0.12	0.70 (0.40, 1.21)	0.20
Tertile1	78.5	Reference		Reference		Reference		Reference	
Tertile 2	79.4	0.86 (0.51, 1.45)	0.56	0.87 (0.51, 1.48)	0.61	0.96 (0.55, 1.64)	0.87	1.01 (0.58, 1.76)	0.96
Tertile 3	91.1	0.31 (0.15, 0.63)	0.001	0.33 (0.16, 0.69)	0.003	0.37 (0.17, 0.78)	0.009	0.41 (0.19, 0.88)	0.02
GSTP1									
Continuous	83.0	0.96 (0.76, 1.21)	0.73	0.93 (0.73, 1.17)	0.52	0.82 (0.62, 1.08)	0.16	0.80 (0.61, 1.05)	0.10
<median	83.8	Reference		Reference		Reference		Reference	
≥median	82.2	1.20 (0.74, 1.95)	0.45	1.13 (0.69, 1.85)	0.63	1.03 (0.62, 1.71)	0.90	0.97 (0.58, 1.61)	0.90
Tertile1	84.0	Reference		Reference		Reference		Reference	
Tertile 2	80.4	1.40 (0.78, 2.53)	0.26	1.36 (0.75, 2.48)	0.32	1.03 (0.52, 2.04)	0.93	1.01 (0.52, 1.99)	0.97
Tertile 3		1.06 (0.57, 1.99)	0.85	0.98 (0.52, 1.86)	0.95	0.80 (0.40, 1.59)	0.52	0.70 (0.35, 1.42)	0.33
GSTA1									
Not expressed	83.4	Reference		Reference		Reference		Reference	
Expressed	79.8	1.41 (0.67, 2.96)	0.37	1.28 (0.60, 2.72)	0.53	1.12 (0.52, 2.41)	0.77	1.09 (0.51, 2.33)	0.83
GSTM1									
Not expressed	83.4	Reference		Reference		Reference		Reference	
Expressed	82.4	0.98 (0.92, 1.05)	0.64	1.11 (0.68, 1.82)	0.67	1.13 (0.69, 1.84)	0.64	1.14 (0.70, 1.88)	0.59
GSTT1									
Not expressed	81.4	Reference		Reference		Reference		Reference	
Expressed	84.9	0.98 (0.91, 1.05)	0.51	0.88 (0.54, 1.43)	0.60	0.90 (0.55, 1.47)	0.68	0.97 (0.59, 1.58)	0.89

	Overall Survival								
ALDH1A1									
Continuous	85.0	0.89 (0.81, 0.98)	0.02	0.91 (0.82, 1.00)	0.05	0.92 (0.83, 1.02)	0.10	0.95 (0.86, 1.06)	0.40
<median	82.5	Reference		Reference		Reference		Reference	
≥median	87.5	0.65 (0.41, 1.04)	0.07	0.74 (0.46, 1.18)	0.21	0.80 (0.49, 1.30)	0.37	0.87 (0.52, 1.43)	0.57
Tertile1	80.0	Reference		Reference		Reference		Reference	
Tertile 2	84.5	0.70 (0.41, 1.17)	0.17	0.72 (0.43, 1.22)	0.23	0.77 (0.45, 1.32)	0.35	0.83 (0.48, 1.43)	0.50
Tertile 3	90.5	0.45 (0.25, 0.80)	0.007	0.52 (0.28, 0.94)	0.03	0.56 (0.30, 1.04)	0.06	0.63 (0.33, 1.19)	0.15
GSTP1									
Continuous	85.0	1.07 (0.85, 1.33)	0.57	1.04 (0.83, 1.30)	0.73	0.97 (0.76, 1.25)	0.83	0.94 (0.74, 1.20)	0.63
<median	86.5	Reference		Reference		Reference		Reference	
≥median	83.5	1.30 (0.83, 2.04)	0.26	1.23 (0.78, 1.96)	0.37	1.14 (0.71, 1.84)	0.58	1.08 (0.67, 1.74)	0.77
Tertile1	86.2	Reference		Reference		Reference		Reference	
Tertile 2	85.2	1.16 (0.66, 2.04)	0.60	1.17 (0.66, 2.06)	0.59	0.93 (0.49, 1.78)	0.83	0.92 (0.49, 1.76)	0.81
Tertile 3	83.6	1.20 (0.68, 2.09)	0.53	1.13 (0.64, 1.99)	0.68	0.95 (0.51, 1.77)	0.87	0.85 (0.45, 1.59)	0.61
GSTA1									
Not expressed	85.2	Reference		Reference		Reference		Reference	
Expressed	83.1	1.20 (0.57, 2.50)	0.63	1.06 (0.50, 2.22)	0.89	0.95 (0.45, 2.01)	0.89	0.89 (0.42, 1.88)	0.76
GSTM1									
Not expressed	86.5	Reference		Reference		Reference		Reference	
Expressed	82.5	1.16 (0.73, 1.84)	0.52	1.20 (0.76, 1.91)	0.43	1.21 (0.76, 1.92)	0.42	1.23 (0.77, 1.95)	0.39
GSTT1									
Not expressed	84.7	Reference		Reference		Reference		Reference	
Expressed	85.3	1.13 (0.72, 1.78)	0.58	1.11 (0.70, 1.74)	0.66	1.13 (0.72, 1.78)	0.60	1.24 (0.78, 1.96)	0.37

[1]Unadjusted, mean(se), DFS or OS, as appropriate
[2]Adjusted for age at diagnosis
[3]Adjusted for age at diagnosis and tumor grade
[4]Adjusted for age at diagnosis, tumor grade, and basal-like subtype
[5]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

Table 30: Disease-Free and Overall Survival by Expression of Genes Involved in Cyclophosphamide Metabolism Among TNBC Participants in the SBCSS Cohort Stratified By Whether Cyclophosphamide was Taken

	Disease-Free Survival					Overall Survival				
	Cyclophosphamide Taken (N=296)		Cyclophosphamide Not Taken (N=122)			Cyclophosphamide Taken (N=296)		Cyclophosphamide Not Taken (N=122)		
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	P_{int}	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	P_{int}
ALDH1A1										
Continuous	1.01 (0.87, 1.17)	0.92	0.91 (0.77, 1.08)	0.30	0.60	1.02 (0.88, 1.17)	0.82	0.88 (0.75, 1.04)	0.12	0.51
<median	Reference		Reference			Reference		Reference		
≥median	0.60 (0.30, 1.20)	0.15	0.92 (0.38, 2.20)	0.85	0.63	0.81 (0.43, 1.52)	0.51	1.01 (0.45, 2.25)	0.99	0.59
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	0.88 (0.43, 1.81)	0.73	1.14 (0.47, 2.74)	0.77	0.65	0.82 (0.41, 1.65)	0.58	0.77 (0.32, 1.83)	0.55	0.88
Tertile 3	0.50 (0.20, 1.25)	0.14	0.30 (0.07, 1.20)	0.09		0.72 (0.33, 1.59)	0.42	0.58 (0.20, 1.70)	0.32	
GSTP1										
Continuous	0.65 (0.47, 0.91)	0.01	1.08 (0.65, 1.80)	0.75	0.10	0.84 (0.62, 1.12)	0.23	1.06 (0.66, 1.70)	0.82	0.41
<median	Reference		Reference			Reference		Reference		
≥median	0.73 (0.37, 1.44)	0.37	1.33 (0.57, 3.12)	0.51	0.28	0.87 (0.47, 1.61)	0.67	1.25 (0.55, 2.84)	0.59	0.50
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	1.33 (0.59, 2.99)	0.49	0.38 (0.10, 1.39)	0.14	0.95	1.09 (0.50, 2.38)	0.82	0.39 (0.11, 1.37)	0.14	0.66
Tertile 3	0.64 (0.25, 1.61)	0.35	0.44 (0.13, 1.51)	0.19		0.85 (0.39, 1.86)	0.68	0.47 (0.14, 1.55)	0.22	
GSTA1										
Not expressed	Reference		Reference			Reference		Reference		
Expressed	0.69 (0.21, 2.29)	0.54	1.64 (0.59, 4.61)	0.35	0.22	0.59 (0.18, 1.93)	0.38	1.25 (0.45, 3.48)	0.66	0.36
GSTM1										
Not expressed	Reference		Reference			Reference		Reference		
Expressed	1.30 (0.71, 2.41)	0.40	0.97 (0.41, 2.31)	0.94	0.50	1.42 (0.80, 2.52)	0.23	0.95 (0.42, 2.15)	0.91	0.42
GSTT1										
Not expressed	Reference		Reference			Reference		Reference		
Expressed	0.99 (0.52, 1.88)	0.98	0.92 (0.41, 2.08)	0.85	1.00	1.44 (0.79, 2.64)	0.23	0.99 (0.46, 2.11)	0.97	0.59

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

The association between individual gene expression levels and the number of cycles of cyclophosphamide completed was evaluated as a surrogate for toxicity (Table 31). No significant associations were observed.

Table 31: Association Between Cyclophosphamide Gene Expression Levels and Cycles of Cyclophosphamide Completed

	OR (95% CI)[1]	***P***
ALDH1A1	0.82 (0.44, 1.52)	0.53
GSTP1	0.66 (0.36, 1.24)	0.20
GSTA1	2.23 (0.62, 8.04)	0.22
GSTM1	1.22 (0.66, 2.23)	0.53
GSTT1	0.87 (0.48, 1.57)	0.64

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

The association between expression of cyclophosphamide metabolizing genes and breast cancer survival among those who took cyclophosphamide was further evaluated through additional adjustment for the number of cycles of cyclophosphamide completed (Table 32). The observed associations between cyclophosphamide gene expression levels and survival outcomes were not materially changed after adjustment for cycles completed.

Table 32: Disease-Free and Overall Survival by Expression of Cyclophosphamide Metabolizing Genes Further Adjusted for Number of Cycles of Cyclophosphamide Completed

	Disease-Free Survival		Overall Survival	
	HR (95% CI)[1]	***P***	**HR (95% CI)[1]**	***P***
ALDH1A1				
Continuous	1.01 (0.87, 1.17)	0.94	1.01 (0.88, 1.17)	0.84
<median	Reference		Reference	
≥median	0.60 (0.30, 1.20)	0.15	0.81 (0.43, 1.52)	0.51
Tertile 1	Reference		Reference	
Tertile 2	0.87 (0.42, 1.80)	0.71	0.82 (0.41, 1.65)	0.57
Tertile 3	0.49 (0.19, 1.22)	0.13	0.71 (0.32, 1.57)	0.40
GSTP1				
Continuous	0.65 (0.47, 0.91)	0.01	0.84 (0.63, 1.12)	0.24
<median	Reference		Reference	
≥median	0.74 (0.38, 1.45)	0.38	0.88 (0.48, 1.64)	0.69
Tertile 1	Reference		Reference	
Tertile 2	1.33 (0.59, 2.99)	0.49	1.10 (0.50, 2.39)	0.82
Tertile 3	0.64 (0.26, 1.63)	0.35	0.86 (0.39, 1.89)	0.71
GSTA1				
Not expressed	Reference		Reference	
Expressed	0.69 (0.21, 2.28)	0.54	0.59 (0.18, 1.92)	0.38
GSTM1				
Not expressed	Reference		Reference	
Expressed	1.32 (0.71, 2.44)	0.38	1.44 (0.81, 2.56)	0.22
GSTT1				
Not expressed	Reference		Reference	
Expressed	0.99 (0.52, 1.88)	0.98	1.45 (0.80, 2.65)	0.22

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, DUSP4 expression, and number of cycles of cyclophosphamide

We stratified our results by timing of survival events, those that occurred in the first three years of follow-up and those that occurred after 3 years, among all women (Table 33) and restricted to those who took cyclophosphamide (Table 34). GSTP1 expression was more strongly associated with better DFS in the first 3 years following diagnosis, particularly in those who took cyclophosphamide (HR=0.45, 95% CI: 0.29, 0.70); no association was observed among those with later events (HR=1.14, 95% CI: 0.67, 1.94). GSTT1 expression was significantly associated with worse OS in the first 3 years following diagnosis among those who took cyclophosphamide (HR=2.76, 95% CI: 1.07, 7.11); no association was observed for OS for events occurring after 3 years (HR=0.92, 95% CI: 0.40, 2.12). A similar pattern was seen for GSTM1 expression although the point estimates were not significant.

Table 33: Association Between Cyclophosphamide Gene Expression Levels Stratified by Early vs. Late Events Among All Participants

	Events <3 years				Events ≥3 years			
	Disease-Free Survival (47 events)		Overall Survival (34 events)		Disease-Free Survival (20 events)		Overall Survival (42 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
ALDH1A1								
Continuous	0.98 (0.85, 1.12)	0.72	1.10 (0.93, 1.32)	0.27	0.94 (0.76, 1.16)	0.56	0.86 (0.75, 0.98)	0.03
<median	Reference		Reference		Reference		Reference	
≥median	0.93 (0.49, 1.78)	0.83	1.22 (0.57, 2.61)	0.61	0.34 (0.12, 0.97)	0.04	0.65 (0.34, 1.27)	0.21
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	1.19 (0.63, 2.27)	0.59	1.16 (0.53, 2.58)	0.71	0.65 (0.22, 1.92)	0.43	0.63 (0.29, 1.33)	0.22
Tertile 3	0.36 (0.13, 0.96)	0.04	0.82 (0.30, 2.24)	0.70	0.47 (0.14, 1.60)	0.23	0.50 (0.22, 1.11)	0.09
GSTP1								
Continuous	0.70 (0.50, 0.97)	0.03	0.86 (0.59, 1.25)	0.43	1.03 (0.64, 1.67)	0.89	1.01 (0.73, 1.41)	0.93
<median	Reference		Reference		Reference		Reference	
≥median	0.81 (0.44, 1.49)	0.50	1.04 (0.50, 2.17)	0.91	1.46 (0.57, 3.77)	0.43	1.07 (0.57, 2.03)	0.83
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.75 (0.33, 1.67)	0.48	0.86 (0.32, 2.34)	0.77	2.09 (0.60, 7.26)	0.25	0.96 (0.41, 2.23)	0.92
Tertile 3	0.55 (0.24, 1.28)	0.17	0.86 (0.32, 2.31)	0.77	1.19 (0.32, 4.36)	0.80	0.82 (0.36, 1.90)	0.65
GSTA1								
Not Expressed	Reference		Reference		Reference		Reference	
Expressed	1.31 (0.57, 2.99)	0.53	0.87 (0.30, 2.53)	0.80	0.51 (0.07, 3.96)	0.52	0.96 (0.33, 2.77)	0.94
GSTM1								
Not Expressed	Reference		Reference		Reference		Reference	
Expressed	1.47 (0.82, 2.63)	0.19	1.67 (0.84, 3.30)	0.14	0.58 (0.21, 1.62)	0.30	0.94 (0.49, 1.80)	0.86
GSTT1								
Not Expressed	Reference		Reference		Reference		Reference	
Expressed	0.94 (0.52, 1.70)	0.83	2.18 (1.08, 4.42)	0.03	1.05 (0.42, 2.59)	0.92	0.79 (0.42, 1.51)	0.48

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

Table 34: Association Between Cyclophosphamide Gene Expression Levels Stratified by Early vs. Late Events Among Those who Took Cyclophosphamide

	Events <3 years				Events ≥3 years			
	Disease-Free Survival (27 events)		Overall Survival (22 events)		Disease-Free Survival (15 events)		Overall Survival (26 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
ALDH1A1								
Continuous	1.01 (0.84, 1.21)	0.90	1.15 (0.96, 1.40)	0.14	1.00 (0.76, 1.30)	0.99	0.90 (0.75, 1.08)	0.26
<median	Reference		Reference		Reference		Reference	
≥median	0.78 (0.33, 1.88)	0.58	1.18 (0.46, 3.08)	0.73	0.36 (0.11, 1.17)	0.09	0.57 (0.24, 1.34)	0.20
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.97 (0.40, 2.35)	0.95	1.52 (0.56, 4.14)	0.41	0.71 (0.20, 2.49)	0.59	0.45 (0.16, 1.26)	0.13
Tertile 3	0.46 (0.13, 1.55)	0.21	0.91 (0.25, 3.32)	0.89	0.50 (0.12, 2.08)	0.34	0.53 (0.19, 1.46)	0.22
GSTP1								
Continuous	0.45 (0.29, 0.70)	0.0004	0.68 (0.45, 1.03)	0.07	1.14 (0.67, 1.94)	0.63	1.02 (0.67, 1.54)	0.93
<median	Reference		Reference		Reference		Reference	
≥median	0.48 (0.20, 1.12)	0.09	0.79 (0.31, 2.03)	0.62	1.56 (0.52, 4.73)	0.43	0.84 (0.36, 1.93)	0.68
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.78 (0.29, 2.10)	0.62	0.79 (0.25, 2.55)	0.70	3.38 (0.87, 13.18)	0.08	1.25 (0.44, 3.52)	0.67
Tertile 3	0.35 (0.11, 1.14)	0.08	0.57 (0.17, 1.98)	0.38	1.66 (0.37, 7.44)	0.51	1.07 (0.38, 3.03)	0.89
GSTA1								
Not Expressed	Reference		Reference		Reference		Reference	
Expressed	0.61 (0.14, 2.65)	0.51	0.37 (0.05, 2.82)	0.34	0.99 (0.12, 8.34)	1.00	0.82 (0.19, 3.55)	0.79
GSTM1								
Not Expressed	Reference		Reference		Reference		Reference	
Expressed	1.90 (0.89, 4.06)	0.10	1.93 (0.83, 4.51)	0.13	0.62 (0.20, 1.97)	0.42	1.11 (0.50, 2.45)	0.80
GSTT1								
Not Expressed	Reference		Reference		Reference		Reference	
Expressed	1.00 (0.45, 2.22)	0.99	2.76 (1.07, 7.11)	0.04	1.02 (0.35, 2.98)	0.97	0.92 (0.40, 2.12)	0.85

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

As previously discussed, the gene expression score based on median cut points of expression was created and the association with clinical and treatment factors was evaluated (Table 35). None of the demographic or clinical factors were associated with the cyclophosphamide gene expression score.

Table 35: Clinical and Treatment Factors by the Cyclophosphamide Gene Score Among TNBC Participants in the SBCSS Cohort

		Median Score	
	N	**mean (SD)**	***P***
Age			
<40	29	2.0 (0.9)	
40-49	154	1.8 (0.8)	0.66
50-59	108	1.9 (0.8)	
≥60	127	1.9 (0.8)	
TNM Stage			
0-I	137	1.9 (0.8)	
IIA	145	1.9 (0.8)	0.83
IIB	85	1.9 (0.8)	
III-IV	38	1.9 (0.8)	
Missing	13	1.8 (0.8)	
Grade			
1	50	1.8 (0.8)	0.79
2	132	1.9 (0.8)	
3	236	1.9 (0.8)	
Chemotherapy			
Yes	390	1.9 (0.8)	0.35
No	28	2.0 (0.8)	
Cyclophosphamide			
Yes	296	1.9 (0.8)	0.26
No	122	2.0 (0.8)	
Radiotherapy			
Yes	103	1.9 (0.8)	0.71
No	315	1.9 (0.8)	
Mastectomy			
Yes	399	1.9 (0.8)	0.75
No	19	1.8 (0.8)	

No association was observed between the additive cyclophosphamide gene expression score and DFS or OS (Table 36). When the association between the cyclophosphamide gene expression score and survival was stratified by whether cyclophosphamide was taken, no significant interaction was observed (Table 37).

Table 36: Disease-Free Survival and Overall Survival by Additive Cyclophosphamide Gene Expression Score Among TNBC Participants in the SBCSS Cohort

	N	HR (95% CI)[1]	*P*	HR (95% CI)[2]	*P*	HR (95% CI)[3]	*P*	HR (95% CI)[4]	*P*
				Disease-Free Survival					
Score[5]									
Group 1	155	Reference		Reference		Reference		Reference	
Group 2	150	1.25 (0.72, 2.16)	0.43	1.25 (0.72, 2.16)	0.43	1.22 (0.70, 2.11)	0.48	1.37 (0.78, 2.41)	0.27
Group 3	113	0.89 (0.47, 1.71)	0.73	0.89 (0.47, 1.71)	0.73	0.90 (0.47, 1.73)	0.76	0.97 (0.50, 1.87)	0.93
P_{trend}		0.83		0.81		0.85		0.96	
				Overall Survival					
Score[5]									
Group 1	155	Reference		Reference		Reference		Reference	
Group 2	150	1.29 (0.76, 2.18)	0.35	1.29 (0.76, 2.18)	0.35	1.24 (0.73, 2.11)	0.42	1.39 (0.81, 2.39)	0.24
Group 3	113	1.17 (0.65, 2.10)	0.61	1.17 (0.65, 2.10)	0.61	1.17 (0.65, 2.10)	0.60	1.26 (0.69, 2.28)	0.45
P_{trend}		0.56		0.55		0.57		0.41	

[1]Adjusted for age at diagnosis
[2]Adjusted for age at diagnosis and tumor grade
[3]Adjusted for age at diagnosis, tumor grade, and basal-like subtype
[4]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression
[5]Score based on median cut point (higher score indicates faster metabolizer)

Table 37: Disease-Free Survival and Overall Survival by Additive Cyclophosphamide Gene Expression Score Among TNBC Participants in the SBCSS Cohort Stratified by Whether Cyclophosphamide was Taken

	Cyclophosphamide Taken		**Cyclophosphamide Not Taken**	
	HR (95% CI)[1]	***P***	**HR (95% CI)[1]**	***P***
		Disease-Free Survival		
Score[2]				
Group 1	Reference		Reference	
Group 2	1.38 (0.70, 2.74)	0.36	1.41 (0.52, 3.87)	0.50
Group 3	0.67 (0.27, 1.65)	0.38	1.57 (0.55, 4.45)	0.40
P_{trend}	0.54		0.40	
		Overall Survival		
Score[2]				
Group 1	Reference		Reference	
Group 2	1.54 (0.79, 3.00)	0.21	1.09 (0.43, 2.81)	0.85
Group 3	1.07 (0.48, 2.38)	0.86	1.38 (0.54, 3.54)	0.51
P_{trend}	0.72		0.50	

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression
[2]Score based on median cut point (higher score indicates faster metabolizer)

The association between the cyclophosphamide gene expression score and the number of cycles of cyclophosphamide completed was evaluated as a surrogate for toxicity. No association was observed (HR=0.90, 95% CI: 0.62, 1.30, p=0.57).

The association between the cyclophosphamide gene expression score and breast cancer survival among those who took cyclophosphamide was further evaluated through additional adjustment for the number of cycles of cyclophosphamide taken (Table 38). The observed associations between the cyclophosphamide gene expression score and survival outcomes were not materially changed after adjustment for cycles completed.

Table 38: Disease-Free and Overall Survival by Additive Cyclophosphamide Gene Score Further Adjusted for Number of Cycles of Cyclophosphamide Completed

	Disease-Free Survival		**Overall Survival**	
	HR (95% CI)[1]	***P***	**HR (95% CI)[1]**	***P***
Score[2]				
Group 1	Reference		Reference	
Group 2	1.41 (0.70, 2.81)	0.33	1.58 (0.81, 3.10)	0.18
Group 3	0.67 (0.27, 1.65)	0.39	1.08 (0.49, 2.40)	0.85

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, DUSP4 expression, and number of cycles of cyclophosphamide
[2]Score based on median cut point (higher score indicates faster metabolizer)

We stratified our results by timing of survival events, those that occurred in the first three years of follow-up and those that occurred after 3 years, among all women and restricted to those who took cyclophosphamide (Table 39). Among all participants, there was a significant difference in the gene expression score for OS between those who had events in the first 3 years compared to those who had later events. The gene expression score was associated with an increased risk of death in the first 3 years (highest score compared to lowest HR=2.68, 95% CI: 1.08, 1.69, P_{trend}=0.03); a similar association was observed when restricted to only those who took cyclophosphamide, although the P_{trend} was no longer significant, potentially due to decreased sample size (HR=2.17, 95% CI: 0.74, 6.33, P_{trend}=0.16). No association was observed between the cyclophosphamide gene expression score and OS for events occurring after 3 years. No significant associations were observed for DFS.

Table 39: Association Between Disease-Free and Overall Survival and Additive Cyclophosphamide Gene Score Stratified by Early vs. Late Events Among All Participants and Only Those who Took Cyclophosphamide

	Events <3 years				Events ≥3 years			
	Disease-Free Survival (47 events)		Overall Survival (34 events)		Disease-Free Survival (20 events)		Overall Survival (42 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
All Participants								
Score[2]								
Group 1	Reference		Reference		Reference		Reference	
Group 2	1.42 (0.71, 2.86)	0.32	2.03 (0.81, 5.10)	0.13	1.26 (0.49, 3.26)	0.63	1.14 (0.57, 2.25)	0.71
Group 3	1.31 (0.62, 2.78)	0.49	2.68 (1.08, 6.65)	0.03	0.36 (0.08, 1.69)	0.19	0.64 (0.26, 1.54)	0.32
P_{trend}	0.46		0.03		0.28		0.40	
Among Those who Took Cyclophosphamide								
Score[2]								
Group 1	Reference		Reference		Reference		Reference	
Group 2	1.11 (0.46, 2.70)	0.82	1.41 (0.47, 4.27)	0.54	1.92 (0.63, 5.90)	0.25	1.57 (0.68, 3.66)	0.29
Group 3	0.85 (0.31, 2.34)	0.75	2.17 (0.74, 6.33)	0.16	0.31 (0.04, 2.69)	0.29	0.34 (0.07, 1.56)	0.17
P_{trend}	0.80		0.16		0.53		0.34	

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression
[2]Score based on median cut point (higher score indicates faster metabolizer)

5-Fluorouracil

Higher tumor grade was associated with expression levels of DPYD and MTHFR below the median and expression levels of TYMS and TYMP above the median (Table 40). Expression levels of TYMS, TYMP, and UMPS above the median were associated with basal-like tumor markers while expression levels of MTHFR below the median were associated with basal-like tumor markers.

Table 40: Clinical and Treatment Factors by Expression of 5-Fluorouracil Metabolizing Genes Among TNBC Participants in the SBCSS Cohort

	DPYD			MTHFR			TYMS			TYMP			UMPS		
	<Median N(%[1])	**≥Median N(%[1])**	***P***	**<Median N(%[1])**	**≥Median N(%[1])**	***P***	**<Median N(%[1])**	**≥Median N(%[1])**	***P***	**<Median N(%[1])**	**≥Median N(%[1])**	***P***	**<Median N(%[1])**	**≥Median N(%[1])**	***P***
N	210	208		213	205		209	209		209	209		211	207	
TNM Stage			0.47			0.42			0.10			0.14			0.78
0-I	65 (47.5)	72 (52.6)		66 (48.2)	71 (51.8)		72 (52.6)	65 (47.5)		68 (49.6)	69 (50.4)		72 (52.6)	65 (47.5)	
IIA	71 (49.0)	74 (51.0)		71 (49.0)	74 (51.0)		74 (51.0)	71 (49.0)		65 (44.8)	80 (55.2)		73 (50.3)	72 (49.7)	
IIB	49 (57.7)	36 (42.4)		50 (58.8)	35 (41.2)		33 (38.8)	52 (61.2)		44 (51.8)	41 (48.2)		40 (47.1)	45 (52.9)	
III-IV	18 (47.4)	20 (52.6)		20 (52.6)	18 (47.4)		23 (60.5)	15 (39.5)		25 (65.8)	13 (34.2)		17 (44.7)	21 (55.3)	
Missing	7	6		6	7		7	6		7	6				
Grade			0.01			<.0001			<.0001			0.0007			0.52
1	16 (32.0)	34 (68.0)		11 (22.0)	39 (78.0)		41 (82.0)	9 (18.0)		37 (74.0)	13 (26.0)		25 (50.0)	25 (50.0)	
2	65 (49.2)	67 (50.8)		54 (40.9)	78 (59.1)		90 (68.2)	42 (31.8)		67 (50.8)	65 (49.2)		72 (54.6)	60 (45.5)	
3	129 (54.7)	107 (45.3)		148 (62.7)	88 (37.3)		78 (33.1)	158 (67.0)		105 (44.5)	131 (55.5)		114 (48.3)	122 (51.7)	
Missing															
Chemotherapy	197 (50.5)	193 (49.5)	0.68	200 (51.3)	190 (48.7)	0.62	191 (49.0)	199 (51.0)	0.12	193 (49.5)	197 (50.5)	0.43	193 (49.5)	197 (50.5)	0.13
5-fluorouracil	160 (50.6)	156 (49.4)	0.78	157 (49.7)	159 (50.3)	0.36	157 (49.7)	159 (50.3)	0.82	163 (51.6)	153 (48.4)	0.25	160 (50.6)	156 (49.4)	
Radiotherapy	59 (57.3)	44 (42.7)	0.10	57 (55.3)	46 (44.7)	0.31	46 (44.7)	57 (55.3)	0.21	57 (55.3)	46 (44.7)	0.21	45 (43.7)	58 (56.3)	0.11
Mastectomy	203 (50.9)	196 (49.1)	0.23	207 (51.9)	192 (48.1)	0.08	198 (49.6)	201 (50.4)	0.48	201 (50.4)	198 (49.6)	0.48	201 (50.4)	198 (49.6)	0.85
Subtype Classification			0.13			<.0001			<.0001			0.0006			0.001
Basal-like	97 (55.8)	77 (44.3)		126 (72.4)	48 (27.6)		23 (13.2)	151 (86.8)		69 (69.7)	105 (60.3)		81 (46.6)	93 (53.5)	
Her-2 Enriched	28 (46.7)	32 (53.3)		23 (38.3)	37 (61.7)		40 (66.7)	20 (33.3)		27 (45.0)	33 (55.0)		38 (63.3)	22 (36.7)	
Luminal A	48 (43.6)	62 (56.4)		29 (26.4)	81 (73.6)		99 (90.0)	11 (10.0)		65 (59.1)	45 (40.9)		59 (53.6)	51 (46.4)	
Luminal B	24 (58.5)	17 (41.5)		22 (53.7)	19 (46.3)		20 (48.8)	21 (51.2)		24 (58.5)	17 (41.5)		11 (26.8)	30 (73.2)	
Normal	13 (39.4)	20 (60.6)		13 (39.4)	20 (60.6)		27 (81.8)	6 (18.2)		24 (72.7)	9 (27.3)		22 (66.7)	11 (33.3)	

[1]Percents shown are row percentages

Median Values: DPYD - 7.877; MTHFR - 6.794; TYMS - 8.785; TYMP - 10.690; UMPS - 6.150

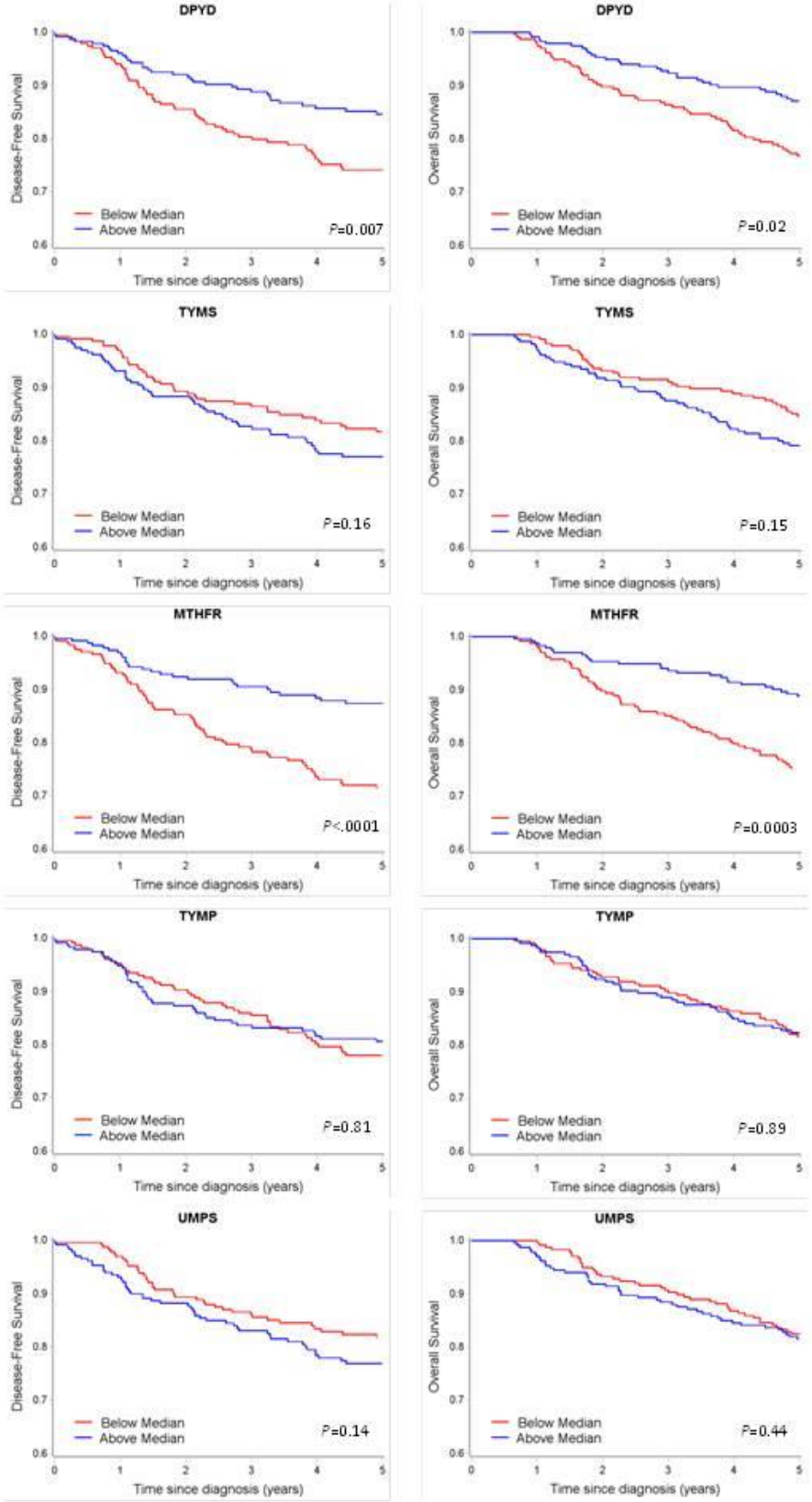
We examined the correlation between the 5-fluorouracil metabolizing genes being investigated in this study and DUSP4 expression (Table 41). DPYD and MTHFR expression was positively correlated with DUSP4 expression (p<.0001 for both) and TYMS expression was negatively associated with DUSP4 expression (p<.0001). No correlation between DUSP4 expression and TYMP or UMPS expression was observed.

Table 41: Correlation between Expression of 5-Fluorouracil Metabolizing Genes and DUSP4 Expression Among TNBC Participants in the SBCSS Cohort

	Correlation with DUSP4	
	r	***p***
DPYD	0.28	<.0001
TYMS	-0.26	<.0001
MTHFR	0.34	<.0001
TYMP	-0.02	0.68
UMPS	0.06	0.23

Univariate analyses between survival, both disease-free and overall, and expression level of genes involved in 5-fluorouracil among all TNBC participants are shown in Figure 3. DPYD expression above the median was significantly associated with better DFS (p=0.007) and better OS (p=0.02) compared with expression levels below the median. MTHFR expression above the median was significantly associated with better DFS (p<.0001) and better OS (p=0.0003) compared with expression levels below the median. No difference was observed in DFS or OS by expression of TYMS, TYMP, or UMPS.

Figure 3: Kaplan-Meier Curves For Disease-Free Survival and Overall Survival for Genes Involved in 5-Fluorouracil Metabolism Among TNBC Participants in the SBCSS Cohort



Among all TNBC participants, DPYD expression above the median was significantly associated with better DFS, compared to the lower median, after adjustment for age at diagnosis and tumor grade (DFS HR=0.59, 95% CI: 0.36, 0.97) (Table 42). Further adjustment for markers of the basal-like subtype and DUSP4 expression slightly attenuated the association but it remained marginally significant for DFS (HR=0.64, 95% CI: 0.38, 1.06). Similar results were observed for analysis by tertile distribution.

TYMS expression was marginally significantly associated with worse OS when analyzed using median or tertile cut points after adjustment for age at diagnosis (HR for upper median compared to lower median=1.55, 95% CI: 0.97, 2.45; HR for tertile 3 compared to tertile 1=1.73, 95% CI: 0.93, 3.02). After further adjustment for tumor grade, basal-like subtype, and DUSP4 expression, the association was completely attenuated.

MTHFR expression above the median was significantly associated with better DFS, compared to the lower median, after adjustment for age at diagnosis and tumor grade (DFS HR=0.54, 95% CI: 0.32, 0.92). Further adjustment for markers of the basal-like subtype and DUSP4 expression attenuated the association (DFS HR=0.65, 95% CI: 0.37, 1.13). A similar pattern was observed when MTHFR expression was analyzed by tertiles.

No association between TYMP expression and DFS or OS was observed.

UMPS expression above the median, compared with below the median, was significantly associated with DFS after adjustment for age at diagnosis and tumor grade (HR=1.81, 95% CI: 1.10, 2.98) and this association was strengthened slightly by adjustment for markers of the basal-

like subtype and DUSP4 expression (HR=1.92, 95% CI: 1.16, 3.19). No association with OS was observed. A similar pattern was observed when UMPS expression was analyzed by tertiles.

Next, the results were stratified by whether the drug of interest, 5-fluorouracil, was taken (Table 43). The protective association between MTHFR expression and survival was only observed among those who did not take 5-fluorouracil and the interaction was significant for OS (p=0.04). No significant differences in the associations between gene expression and survival outcomes were observed for DPYD, TYMS, TYMP, or UMPS by whether 5-fluorouracil was taken.

Table 42: Disease-Free and Overall Survival by Expression of Genes Involved in 5-Fluorouracil Metabolism Among TNBC Participants in the SBCSS Cohort

	5-yr Rate, %[1]	HR (95% CI)[2]	*P*	HR (95% CI)[3]	*P*	HR (95% CI)[4]	*P*	HR (95% CI)[5]	*P*
				Disease-Free Survival					
DPYD									
Continuous	83.0	0.93 (0.82, 1.05)	0.23	0.93 (0.83, 1.05)	0.25	0.93 (0.82, 1.05)	0.24	1.00 (0.86, 1.16)	0.98
<median	78.4	Reference		Reference		Reference		Reference	
≥median	87.5	0.56 (0.34, 0.92)	0.02	0.59 (0.36, 0.97)	0.04	0.59 (0.36, 0.98)	0.04	0.64 (0.38, 1.06)	0.08
Tertile 1	76.4	Reference		Reference		Reference		Reference	
Tertile 2	86.5	0.51 (0.28, 0.91)	0.02	0.51 (0.28, 0.91)	0.02	0.52 (0.29, 0.93)	0.03	0.58 (0.32, 1.05)	0.07
Tertile 3	85.7	0.57 (0.32, 1.02)	0.06	0.61 (0.34, 1.09)	0.10	0.63 (0.35, 1.13)	0.12	0.70 (0.38, 1.27)	0.24
TYMS									
Continuous	83.0	1.06 (0.93, 1.22)	0.37	1.02 (0.88, 1.17)	0.83	0.96 (0.83, 1.11)	0.58	0.98 (0.86, 1.12)	0.79
<median	85.5	Reference		Reference		Reference		Reference	
≥median	80.5	1.60 (0.98, 2.62)	0.06	1.44 (0.84, 2.47)	0.18	1.14 (0.61, 2.16)	0.68	1.15 (0.63, 2.12)	0.65
Tertile 1	84.3	Reference		Reference		Reference		Reference	
Tertile 2	84.8	1.04 (0.56, 1.95)	0.90	0.94 (0.49, 1.82)	0.86	0.78 (0.38, 1.59)	0.49	0.77 (0.38, 1.55)	0.46
Tertile 3	79.8	1.61 (0.89, 2.90)	0.11	1.36 (0.69, 2.67)	0.37	0.93 (0.41, 2.13)	0.87	0.91 (0.41, 2.00)	0.81
MTHFR									
Continuous	83.0	0.95 (0.84, 1.08)	0.43	0.96 (0.84, 1.09)	0.53	0.97 (0.85, 1.12)	0.69	1.05 (0.90, 1.23)	0.51
<median	77.7	Reference		Reference		Reference		Reference	
≥median	88.4	0.51 (0.31, 0.84)	0.009	0.54 (0.32, 0.92)	0.02	0.59 (0.34, 1.02)	0.06	0.65 (0.37, 1.13)	0.13
Tertile 1	77.4	Reference		Reference		Reference		Reference	
Tertile 2	83.4	0.67 (0.38, 1.17)	0.15	0.68 (0.39, 1.19)	0.17	0.72 (0.41, 1.27)	0.25	0.81 (0.45, 1.45)	0.47
Tertile 3	88.2	0.48 (0.26, 0.89)	0.02	0.53 (0.28, 1.01)	0.05	0.60 (0.31, 1.17)	0.13	0.68 (0.34, 1.36)	0.28
TYMP									
Continuous	83.0	0.98 (0.76, 1.26)	0.85	0.93 (0.72, 1.21)	0.59	0.91 (0.70, 1.18)	0.46	0.93 (0.73, 1.18)	0.55
<median	83.0	Reference		Reference		Reference		Reference	
≥median	82.8	1.16 (0.72, 1.87)	0.55	1.08 (0.67, 1.76)	0.75	1.03 (0.63, 1.68)	0.92	1.05 (0.64, 1.72)	0.84
Tertile 1	81.9	Reference		Reference		Reference		Reference	
Tertile 2	83.2	1.04 (0.58, 1.85)	0.90	1.00 (0.56, 1.79)	1.00	0.95 (0.53, 1.70)	0.86	1.00 (0.55, 1.79)	0.99
Tertile 3	83.7	0.94 (0.52, 1.72)	0.85	0.86 (0.47, 1.58)	0.63	0.82 (0.44, 1.50)	0.51	0.83 (0.45, 1.54)	0.56

UMPS									
Continuous	83.0	1.04 (0.91, 1.18)	0.58	1.02 (0.90, 1.16)	0.78	1.01 (0.89, 1.15)	0.84	1.06 (0.92, 1.22)	0.40
<median	86.8	Reference		Reference		Reference		Reference	
≥median	79.2	1.84 (1.12, 3.03)	0.02	1.81 (1.10, 2.98)	0.02	1.79 (1.09, 2.94)	0.02	1.92 (1.16, 3.19)	0.01
Tertile 1	85.1	Reference		Reference		Reference		Reference	
Tertile 2	86.2	0.96 (0.51, 1.84)	0.91	0.91 (0.47, 1.74)	0.77	0.87 (0.46, 1.67)	0.68	1.00 (0.51, 1.94)	0.99
Tertile 3	77.5	1.79 (1.00, 3.19)	0.048	1.70 (0.95, 3.04)	0.07	1.67 (0.93, 2.98)	0.08	1.96 (1.07, 3.58)	0.03

Overall Survival

DPYD									
Continuous	85.0	0.92 (0.82, 1.04)	0.18	0.93 (0.83, 1.04)	0.22	0.93 (0.82, 1.04)	0.21	1.00 (0.87, 1.15)	1.00
<median	81.2	Reference		Reference		Reference		Reference	
≥median	88.8	0.70 (0.44, 1.10)	0.12	0.74 (0.47, 1.17)	0.20	0.75 (0.47, 1.18)	0.21	0.82 (0.51, 1.32)	0.42
Tertile 1	78.7	Reference		Reference		Reference		Reference	
Tertile 2	89.4	0.62 (0.36, 1.07)	0.09	0.60 (0.35, 1.04)	0.07	0.61 (0.35, 1.06)	0.08	0.69 (0.39, 1.22)	0.21
Tertile 3	86.8	0.70 (0.41, 1.21)	0.20	0.74 (0.43, 1.27)	0.28	0.76 (0.44, 1.31)	0.33	0.88 (0.50, 1.56)	0.67
TYMS									
Continuous	85.0	1.09 (0.95, 1.24)	0.24	1.01 (0.88, 1.16)	0.89	0.96 (0.83, 1.11)	0.57	0.98 (0.86, 1.12)	0.79
<median	88.0	Reference		Reference		Reference		Reference	
≥median	82.1	1.55 (0.97, 2.45)	0.06	1.29 (0.79, 2.11)	0.32	1.05 (0.58, 1.91)	0.86	1.04 (0.59, 1.83)	0.90
Tertile 1	88.4	Reference		Reference		Reference		Reference	
Tertile 2	85.1	1.17 (0.65, 2.11)	0.60	1.03 (0.56, 1.87)	0.94	0.90 (0.47, 1.72)	0.75	0.88 (0.46, 1.68)	0.70
Tertile 3	81.5	1.73 (0.99, 3.02)	0.06	1.37 (0.74, 2.53)	0.32	1.06 (0.49, 2.26)	0.89	1.00 (0.48, 2.07)	0.99
MTHFR									
Continuous	85.0	0.91 (0.81, 1.03)	0.13	0.92 (0.82, 1.04)	0.20	0.93 (0.82, 1.06)	0.27	1.00 (0.87, 1.16)	0.95
<median	81.0	Reference		Reference		Reference		Reference	
≥median	89.2	0.62 (0.39, 0.98)	0.04	0.71 (0.44, 1.14)	0.16	0.77 (0.47, 1.27)	0.31	0.87 (0.52, 1.45)	0.60
Tertile 1	79.1	Reference		Reference		Reference		Reference	
Tertile 2	83.6	0.76 (0.46, 1.27)	0.30	0.78 (0.47, 1.31)	0.35	0.82 (0.49, 1.38)	0.46	0.93 (0.55, 1.59)	0.79
Tertile 3	92.6	0.42 (0.23, 0.76)	0.004	0.49 (0.26, 0.92)	0.03	0.53 (0.28, 1.02)	0.06	0.62 (0.32, 1.21)	0.16

TYMP									
Continuous	85.0	1.02 (0.81, 1.30)	0.85	0.95 (0.75, 1.21)	0.70	0.93 (0.73, 1.19)	0.58	0.96 (0.77, 1.21)	0.75
<median	85.5	Reference		Reference		Reference		Reference	
≥median	84.5	1.18 (0.75, 1.85)	0.48	1.05 (0.67, 1.66)	0.82	1.01 (0.64, 1.60)	0.95	1.04 (0.66, 1.65)	0.85
Tertile 1	86.2	Reference		Reference		Reference		Reference	
Tertile 2	83.1	1.14 (0.67, 1.95)	0.63	1.11 (0.65, 1.90)	0.70	1.06 (0.62, 1.81)	0.84	1.13 (0.66, 1.95)	0.66
Tertile 3	85.8	0.96 (0.54, 1.71)	0.90	0.84 (0.47, 1.50)	0.56	0.81 (0.45, 1.45)	0.47	0.84 (0.47, 1.50)	0.55
UMPS									
Continuous	85.0	0.99 (0.89, 1.11)	0.92	0.97 (0.86, 1.09)	0.59	0.96 (0.86, 1.08)	0.55	1.00 (0.88, 1.13)	0.99
<median	86.5	Reference		Reference		Reference		Reference	
≥median	83.5	1.39 (0.88, 2.20)	0.16	1.38 (0.88, 2.18)	0.16	1.37 (0.87, 2.17)	0.17	1.44 (0.91, 2.28)	0.12
Tertile 1	86.7	Reference		Reference		Reference		Reference	
Tertile 2	84.7	0.94 (0.53, 1.66)	0.83	0.85 (0.48, 1.52)	0.59	0.83 (0.47, 1.48)	0.53	0.93 (0.51, 1.67)	0.80
Tertile 3	83.6	1.31 (0.76, 2.28)	0.33	1.24 (0.71, 2.15)	0.45	1.22 (0.70, 2.11)	0.49	1.38 (0.78, 2.44)	0.26

[1]Unadjusted, mean(se), DFS or OS, as appropriate
[2]Adjusted for age at diagnosis
[3]Adjusted for age at diagnosis and tumor grade
[4]Adjusted for age at diagnosis, tumor grade, and basal-like subtype
[5]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

Table 43: Disease-Free and Overall Survival by Expression of Genes Involved in 5-Fluorouracil Metabolism Among TNBC Participants in the SBCSS Cohort Stratified By Whether 5-Fluorouracil was Taken

	Disease-Free Survival					Overall Survival				
	5-FU Taken (N=316)		5-FU Not Taken (N=102)			5-FU Taken (N=316)		5-FU Not Taken (N=102)		
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	P_{int}	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	P_{int}
DPYD										
Continuous	1.13 (0.94, 1.37)	0.19	0.86 (0.70, 1.06)	0.15	0.39	1.12 (0.93, 1.35)	0.22	0.87 (0.71, 1.08)	0.20	0.78
<median	Reference		Reference			Reference		Reference		
≥median	0.76 (0.40, 1.43)	0.39	0.49 (0.20, 1.18)	0.11	0.64	1.19 (0.66, 2.14)	0.56	0.42 (0.18, 1.00)	0.05	0.19
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	0.61 (0.28, 1.33)	0.21	0.74 (0.27, 1.99)	0.55	0.69	0.79 (0.38, 1.66)	0.54	0.75 (0.29, 1.92)	0.55	0.34
Tertile 3	0.86 (0.41, 1.82)	0.70	0.52 (0.18, 1.49)	0.22		1.33 (0.65, 2.71)	0.43	0.46 (0.16, 1.31)	0.15	
TYMS										
Continuous	1.03 (0.87, 1.24)	0.71	0.96 (0.78, 1.18)	0.71	0.58	1.02 (0.86, 1.22)	0.80	0.98 (0.80, 1.20)	0.85	0.74
<median	Reference		Reference			Reference		Reference		
≥median	1.55 (0.75, 3.19)	0.24	0.45 (0.12, 1.65)	0.23	0.13	1.13 (0.58, 2.22)	0.72	0.69 (0.20, 2.39)	0.56	0.48
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	1.35 (0.57, 3.18)	0.49	0.25 (0.05, 1.28)	0.10	0.16	1.24 (0.58, 2.68)	0.58	0.53 (0.14, 1.98)	0.34	0.60
Tertile 3	1.40 (0.54, 3.64)	0.49	0.31 (0.05, 1.85)	0.20		1.10 (0.46, 2.66)	0.83	0.87 (0.20, 3.75)	0.85	
MTHFR										
Continuous	1.18 (0.96, 1.45)	0.11	0.95 (0.77, 1.19)	0.68	0.92	1.10 (0.90, 1.33)	0.35	0.93 (0.76, 1.14)	0.49	0.80
<median	Reference		Reference			Reference		Reference		
≥median	1.10 (0.56, 2.15)	0.79	0.23 (0.08, 0.71)	0.01	0.08	1.42 (0.76, 2.64)	0.27	0.35 (0.13, 0.91)	0.03	0.12
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	1.31 (0.62, 2.77)	0.48	0.54 (0.19, 1.56)	0.26	0.13	1.30 (0.65, 2.59)	0.46	0.77 (0.31, 1.87)	0.56	0.04
Tertile 3	1.40 (0.58, 3.39)	0.45	0.26 (0.07, 0.94)	0.04		1.48 (0.66, 3.35)	0.34	0.07 (0.01, 0.56)	0.01	

TYMP										
Continuous	0.90 (0.70, 1.18)	0.45	0.86 (0.54, 1.36)	0.51	0.90	0.97 (0.75, 1.24)	0.80	0.79 (0.51, 1.23)	0.30	0.75
<median	Reference		Reference			Reference		Reference		
≥median	0.91 (0.49, 1.68)	0.76	1.14 (0.48, 2.68)	0.77	0.61	0.99 (0.56, 1.76)	0.98	0.91 (0.40, 2.05)	0.82	0.93
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	0.73 (0.35, 1.54)	0.41	1.57 (0.57, 4.34)	0.39	0.90	0.91 (0.46, 1.79)	0.78	1.52 (0.58, 4.01)	0.39	0.99
Tertile 3	0.74 (0.35, 1.55)	0.43	0.78 (0.24, 2.51)	0.68		0.76 (0.38, 1.54)	0.45	0.70 (0.23, 2.15)	0.53	
UMPS										
Continuous	1.12 (0.94, 1.34)	0.21	1.09 (0.85, 1.41)	0.49	0.81	0.98 (0.85, 1.14)	0.84	1.18 (0.90, 1.55)	0.23	0.17
<median	Reference		Reference			Reference		Reference		
≥median	2.24 (1.18, 4.28)	0.01	1.92 (0.80, 4.61)	0.15	0.73	1.42 (0.81, 2.50)	0.22	1.75 (0.76, 4.03)	0.19	0.76
Tertile 1	Reference		Reference			Reference		Reference		
Tertile 2	0.96 (0.41, 2.24)	0.92	1.43 (0.47, 4.35)	0.53	0.71	0.90 (0.44, 1.88)	0.79	1.24 (0.44, 3.51)	0.68	0.50
Tertile 3	2.04 (0.96, 4.33)	0.06	2.32 (0.81, 6.58)	0.12		1.31 (0.65, 2.63)	0.44	1.70 (0.63, 4.58)	0.29	

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

The association between individual gene expression levels and the number of cycles of 5-fluorouracil completed was evaluated as a surrogate for toxicity (Table 44). Participants with TYMP expression levels above the median were significantly less likely to complete 6 or more cycles as compared with those with TYMP expression level below the median (OR=0.40, 95% CI: 0.22, 0.72, p=0.002). No other significant associations were observed.

Table 44: Association Between 5-Fluorouracil Gene Expression Levels and Cycles of 5-Fluorouracil Completed

	OR (95% CI)[1]	***P***
DPYD	0.82 (0.47, 1.42)	0.47
TYMS	1.57 (0.79, 3.14)	0.20
MTHFR	0.63 (0.35, 1.16)	0.14
TYMP	0.40 (0.22, 0.72)	0.002
UMPS	1.12 (0.65, 1.94)	0.69

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

The association between 5-fluorouracil gene expression levels and breast cancer survival among those who took 5-fluorouracil was further evaluated through additional adjustment for the number of cycles of cyclophosphamide taken (Table 45). The observed associations between 5-fluorouracil gene expression levels and survival outcomes were not materially changed after adjustment for cycles completed.

Table 45: Disease-Free and Overall Survival by Expression of 5-Fluorouracil Metabolizing Genes Further Adjusted for Number of Cycles of 5-Fluorouracil Completed

	Disease-Free Survival		Overall Survival	
	HR (95% CI)[1]	***P***	**HR (95% CI)[1]**	***P***
DPYD				
Continuous	1.13 (0.94, 1.35)	0.21	1.12 (0.93, 1.34)	0.23
<median	Reference		Reference	
≥median	0.74 (0.39, 1.40)	0.36	1.18 (0.66, 2.12)	0.58
Tertile 1	Reference		Reference	
Tertile 2	0.62 (0.28, 1.37)	0.24	0.80 (0.38, 1.68)	0.56
Tertile 3	0.86 (0.41, 1.80)	0.68	1.32 (0.65, 2.69)	0.44
TYMS				
Continuous	1.03 (0.86, 1.23)	0.75	1.02 (0.85, 1.21)	0.84
<median	Reference		Reference	
≥median	1.52 (0.74, 3.14)	0.26	1.11 (0.57, 2.18)	0.76
Tertile 1	Reference		Reference	
Tertile 2	1.37 (0.58, 3.23)	0.47	1.25 (0.58, 2.69)	0.57
Tertile 3	1.38 (0.53, 3.57)	0.51	1.09 (0.45, 2.61)	0.85
MTHFR				
Continuous	1.20 (0.97, 1.48)	0.09	1.10 (0.91, 1.34)	0.33
<median	Reference		Reference	
≥median	1.11 (0.56, 2.18)	0.76	1.43 (0.76, 2.67)	0.27
Tertile 1	Reference		Reference	
Tertile 2	1.30 (0.61, 2.75)	0.49	1.30 (0.65, 2.59)	0.46
Tertile 3	1.52 (0.62, 3.77)	0.36	1.55 (0.68, 3.56)	0.30
TYMP				
Continuous	0.90 (0.69, 1.17)	0.43	0.97 (0.75, 1.24)	0.79
<median	Reference		Reference	
≥median	0.93 (0.50, 1.72)	0.81	1.02 (0.57, 1.81)	0.95
Tertile 1	Reference		Reference	
Tertile 2	0.74 (0.35, 1.56)	0.43	0.92 (0.47, 1.83)	0.82
Tertile 3	0.73 (0.35, 1.54)	0.41	0.76 (0.38, 1.55)	0.46
UMPS				
Continuous	1.11 (0.93, 1.33)	0.24	0.98 (0.85, 1.14)	0.80
<median	Reference		Reference	
≥median	2.25 (1.18, 4.32)	0.01	1.42 (0.81, 2.49)	0.23
Tertile 1	Reference		Reference	
Tertile 2	0.99 (0.42, 2.33)	0.97	0.92 (0.44, 1.92)	0.82
Tertile 3	2.08 (0.97, 4.45)	0.06	1.32 (0.66, 2.66)	0.43

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, DUSP4 expression, and number of cycles of 5-fluorouracil

We stratified our results by timing of survival events, those that occurred in the first three years of follow-up and those that occurred after 3 years, among all women (Table 46) and restricted to those who took 5-fluorouracil (Table 47). The effect estimates for TYMS expression and MTHFR expression were greater in magnitude for events that occurred more than 3 years after diagnosis, particularly in those who took 5-fluorouracil, although neither association was significant; no association was observed among those with earlier events. UMPS expression was associated with significantly worse DFS among those who experienced a survival event in the first three years following breast cancer diagnosis (HR=2.01, 95% CI: 1.10, 3.70, p=0.02), particularly among those who were treated with 5-fluorouracil (HR=2.33, 95% CI: 1.05, 5.17, p=0.04); however, while not significant, the point estimate for those who had a later event was not significantly different from the HR associated with early events. No other differences between those who had events in the first 3 years and those after 3 years were apparent.

Table 46: Association Between 5-fluorouracil Gene Expression Levels Stratified by Early vs. Late Events Among All Participants

	Events <3 years				Events ≥3 years			
	Disease-Free Survival (47 events)		Overall Survival (34 events)		Disease-Free Survival (20 events)		Overall Survival (42 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
DPYD								
Continuous	1.01 (0.85, 1.20)	0.94	1.10 (0.89, 1.36)	0.39	0.98 (0.75, 1.29)	0.88	0.92 (0.77, 1.10)	0.37
<median	Reference		Reference		Reference		Reference	
≥median	1.01 (0.85, 1.20)	0.23	0.66 (0.32, 1.38)	0.27	0.52 (0.20, 1.33)	0.17	0.97 (0.52, 1.81)	0.92
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.49 (0.24, 1.04)	0.06	0.97 (0.20, 4.78)	0.97	0.77 (0.27, 2.23)	0.63	0.88 (0.41, 1.90)	0.74
Tertile 3	0.73 (0.36, 1.46)	0.37	1.40 (0.29, 6.69)	0.67	0.61 (0.19, 1.98)	0.41	1.12 (0.51, 2.44)	0.78
TYMS								
Continuous	0.95 (0.81, 1.11)	0.50	0.98 (0.81, 1.18)	0.80	1.07 (0.81, 1.43)	0.62	0.99 (0.81, 1.20)	0.90
<median	Reference		Reference		Reference		Reference	
≥median	1.00 (0.48, 2.07)	0.99	1.03 (0.44, 2.41)	0.95	1.57 (0.53, 4.64)	0.41	1.06 (0.50, 2.27)	0.88
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.94 (0.41, 2.18)	0.89	1.78 (0.64, 4.95)	0.27	0.38 (0.09, 1.60)	0.19	0.48 (0.19, 1.19)	0.11
Tertile 3	0.78 (0.29, 2.05)	0.61	1.20 (0.37, 3.86)	0.76	1.32 (0.34, 5.11)	0.14	0.98 (0.37, 2.60)	0.97
MTHFR								
Continuous	1.02 (0.85, 1.23)	0.82	1.10 (0.88, 1.37)	0.42	1.15 (0.83, 1.59)	0.41	0.95 (0.78, 1.14)	0.56
<median	Reference		Reference		Reference		Reference	
≥median	0.78 (0.40, 1.52)	0.47	0.88 (0.40, 1.92)	0.74	0.42 (0.15, 1.17)	0.10	0.83 (0.42, 1.64)	0.60
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.79 (0.40, 1.57)	0.51	1.02 (0.21, 5.01)	0.98	0.85 (0.28, 2.58)	0.78	0.90 (0.43, 1.87)	0.77
Tertile 3	0.63 (0.27, 1.48)	0.29	1.46 (0.31, 6.98)	0.63	0.79 (0.24, 2.59)	0.70	0.57 (0.24, 1.35)	0.20
TYMP								
Continuous	1.01 (0.75, 1.34)	0.96	0.92 (0.67, 1.26)	0.60	0.77 (0.49, 1.19)	0.24	0.98 (0.71, 1.35)	0.90
<median	Reference		Reference		Reference		Reference	
≥median	1.37 (0.75, 2.49)	0.30	1.10 (0.55, 2.21)	0.78	0.57 (0.23, 1.43)	0.23	0.98 (0.53, 1.80)	0.94
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.84 (0.26, 2.66)	0.76	1.16 (0.51, 2.62)	0.72	0.85 (0.30, 2.39)	0.76	1.13 (0.54, 2.35)	0.75
Tertile 3	0.86 (0.27, 2.74)	0.19	1.00 (0.97, 1.04)	0.50	0.53 (0.17, 1.66)	0.28	0.88 (0.40, 1.93)	0.75

UMPS								
Continuous	1.05 (0.89, 1.23)	0.55	1.11 (0.91, 1.36)	0.31	1.09 (0.84, 1.41)	0.52	0.94 (0.81, 1.09)	0.42
<median	Reference		Reference		Reference		Reference	
≥median	2.01 (1.10, 3.70)	0.02	2.14 (1.03, 4.43)	0.04	1.72 (0.69, 4.28)	0.24	1.10 (0.60, 2.02)	0.77
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	1.06 (0.48, 2.34)	0.89	1.18 (0.46, 3.07)	0.73	0.87 (0.26, 2.95)	0.82	6.27 (0.83, 47.62)	0.08
Tertile 3	1.98 (0.96, 4.10)	0.07	2.07 (0.86, 4.99)	0.10	1.93 (0.65, 5.75)	0.24	4.47 (0.57, 34.79)	0.15

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

Table 47: Association Between 5-fluorouracil Gene Expression Levels Stratified by Early vs. Late Events Among Those who Took 5-Fluorouracil

	Events <3 years				Events ≥3 years			
	Disease-Free Survival (30 events)		Overall Survival (22 events)		Disease-Free Survival (14 events)		Overall Survival (29 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
DPYD								
Continuous	1.21 (0.97, 1.51)	0.10	1.26 (0.98, 1.61)	0.07	0.97 (0.71, 1.32)	0.85	0.98 (0.77, 1.24)	0.86
<median	Reference		Reference		Reference		Reference	
≥median	0.89 (0.41, 1.92)	0.76	1.33 (0.53, 3.32)	0.55	0.55 (0.18, 1.69)	0.30	1.10 (0.51, 2.35)	0.81
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.61 (0.23, 1.61)	0.32	0.72 (0.22, 2.35)	0.59	0.60 (0.16, 2.25)	0.45	0.83 (0.32, 2.15)	0.70
Tertile 3	0.95 (0.38, 2.36)	0.91	1.25 (0.43, 3.64)	0.69	0.71 (0.19, 2.65)	0.61	1.41 (0.54, 3.67)	0.48
TYMS								
Continuous	1.03 (0.84, 1.28)	0.76	0.99 (0.80, 1.23)	0.92	1.05 (0.75, 1.45)	0.79	1.09 (0.82, 1.47)	0.55
<median	Reference		Reference		Reference		Reference	
≥median	1.27 (0.53, 3.04)	0.60	0.92 (0.34, 2.47)	0.86	2.42 (0.68, 8.65)	0.17	1.37 (0.56, 3.30)	0.49
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	1.53 (0.53, 4.42)	0.43	3.12 (0.82, 11.86)	0.10	0.88 (0.18, 4.18)	0.87	0.57 (0.19, 1.69)	0.31
Tertile 3	1.10 (0.33, 3.66)	0.88	1.12 (0.24, 5.22)	0.88	2.57 (0.53, 12.42)	0.24	1.34 (0.43, 4.16)	0.61
MTHFR								
Continuous	1.13 (0.91, 1.42)	0.27	1.24 (0.97, 1.60)	0.09	1.62 (0.81, 3.24)	0.17	0.98 (0.75, 1.29)	0.90
<median	Reference		Reference		Reference		Reference	
≥median	1.35 (0.60, 3.06)	0.47	1.08 (0.40, 2.87)	0.88	0.74 (0.23, 2.36)	0.61	1.68 (0.73, 3.84)	0.22
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	1.19 (0.49, 2.86)	0.70	1.31 (0.46, 3.72)	0.61	1.79 (0.41, 7.90)	0.44	1.34 (0.52, 3.45)	0.55
Tertile 3	1.12 (0.37, 3.43)	0.84	1.51 (0.41, 5.62)	0.53	2.24 (0.46, 10.87)	0.32	1.45 (0.50, 4.19)	0.50
TYMP								
Continuous	0.93 (0.68, 1.27)	0.64	0.86 (0.60, 1.22)	0.39	0.80 (0.47, 1.36)	0.41	1.02 (0.71, 1.47)	0.91
<median	Reference		Reference		Reference		Reference	
≥median	1.08 (0.51, 2.31)	0.84	0.92 (0.37, 2.27)	0.86	0.60 (0.20, 1.80)	0.36	1.00 (0.48, 2.10)	1.00
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.83 (0.33, 2.08)	0.70	0.91 (0.32, 2.54)	0.85	0.59 (0.16, 2.16)	0.42	0.91 (0.37, 2.27)	0.85
Tertile 3	0.84 (0.34, 2.07)	0.70	0.55 (0.18, 1.66)	0.29	0.55 (0.15, 2.06)	0.37	0.88 (0.35, 2.20)	0.78

UMPS								
Continuous	1.17 (0.94, 1.46)	0.16	1.18 (0.91, 1.53)	0.20	1.06 (0.80, 1.42)	0.68	0.91 (0.77, 1.09)	0.31
<median	Reference		Reference		Reference		Reference	
≥median	2.33 (1.05, 5.17)	0.04	2.58 (0.99, 6.70)	0.05	2.19 (0.72, 6.70)	0.17	1.02 (0.49, 2.11)	0.97
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	1.08 (0.38, 3.07)	0.88	1.71 (0.48, 6.15)	0.41	0.77 (0.17, 3.57)	0.74	0.68 (0.27, 1.71)	0.41
Tertile 3	2.03 (0.80, 5.20)	0.14	2.46 (0.75, 8.07)	0.14	2.28 (0.63, 8.30)	0.21	0.96 (0.39, 2.36)	0.94

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

As previously described, a gene expression score based on the median cut point of expression was created and the association with clinical and treatment factors was evaluated (Table 48); higher scores are associated with higher drug exposure. The 5-fluorouracil gene expression score was significantly associated with higher grade (p=0.0007). No other significant associations were observed.

Table 48: Clinical and Treatment Factors by 5-Fluorouracil Gene Expression Score Among TNBC Participants in the SBCSS Cohort

		Median Score	
	N	**mean (SD)**	***P***
Age			
<40	29	2.5 (0.9)	
40-49	154	2.5 (0.9)	0.55
50-59	108	2.6 (1.0)	
≥60	127	2.4 (1.0)	
TNM Stage			
0-I	137	2.5 (1.0)	
IIA	145	2.5 (1.0)	0.74
IIB	85	2.6 (0.9)	
III-IV	38	2.5 (1.0)	
Missing	13	2.6 (1.0)	
Grade			
1	50	2.3 (0.9)	0.0007
2	132	2.3 (0.9)	
3	236	2.7 (1.0)	
Chemotherapy			
Yes	390	2.5 (1.0)	0.58
No	28	2.6 (0.8)	
5-Fluorouracil			
Yes	316	2.5 (1.0)	0.81
No	102	2.5 (0.9)	
Radiotherapy			
Yes	103	2.5 (0.9)	0.86
No	315	2.5 (1.0)	
Mastectomy			
Yes	399	2.5 (1.0)	0.37
No	19	2.3 (0.9)	

No association was observed between the additive 5-fluorouracil gene expression score and DFS or OS (Table 49). The 5-fluorouracil gene expression score was associated with better DFS after adjustment for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression; although the association was not significant (highest score compared to lowest score HR=0.51, 95% CI: 0.20, 1.29, P_{trend}=0.16). Although higher grade is associated with a higher 5-fluorouracil gene expression score, there is no difference in the observed results when grade is not adjusted for.

Table 49: Disease-Free Survival and Overall Survival by Additive 5-Fluorouracil Gene Score Among TNBC Participants in the SBCSS Cohort

	N	HR (95% CI)[1]	*P*	HR (95% CI)[2]	*P*	HR (95% CI)[3]	*P*	HR (95% CI)[4]	*P*
Disease-Free Survival									
Score[5]									
Group 1	67	Reference		Reference		Reference		Reference	
Group 2	142	1.02 (0.50, 2.08)	0.96	1.01 (0.49, 2.07)	0.99	0.89 (0.43, 1.85)	0.75	0.84 (0.40, 1.75)	0.64
Group 3	138	1.05 (0.52, 2.15)	0.89	0.98 (0.48, 2.02)	0.97	0.82 (0.39, 1.72)	0.59	0.79 (0.37, 1.66)	0.53
Group 4	71	0.79 (0.33, 1.90)	0.60	0.72 (0.29, 1.75)	0.47	0.55 (0.22, 1.40)	0.21	0.51 (0.20, 1.29)	0.15
P_{trend}		0.68		0.50		0.21		0.16	
Overall Survival									
Score[5]									
Group 1	67	Reference		Reference		Reference		Reference	
Group 2	142	1.11 (0.56, 2.21)	0.76	1.12 (0.56, 2.23)	0.75	1.01 (0.51, 2.04)	0.97	0.97 (0.48, 1.96)	0.93
Group 3	138	1.12 (0.57, 2.22)	0.74	1.02 (0.51, 2.04)	0.95	0.87 (0.43, 1.79)	0.71	0.85 (0.41, 1.74)	0.65
Group 4	71	1.02 (0.46, 2.27)	0.96	0.91 (0.40, 2.04)	0.81	0.73 (0.31, 1.70)	0.46	0.67 (0.29, 1.57)	0.36
P_{trend}		0.96		0.72		0.37		0.29	

[1]Adjusted for age at diagnosis
[2]Adjusted for age at diagnosis and grade
[3]Adjusted for age at diagnosis, grade, and basal-like-subtype
[4]Adjusted for age at diagnosis, grade, basal-like subtype, and DUSP4
[5]Score based on median cut point (higher score indicates faster metabolizer, slower activator, and poorer response)

When the association between the 5-fluorouracil gene score and survival was stratified by whether 5-fluorouracil was taken, no significant associations were observed for either stratum (Table 50).

Table 50: Disease-Free Survival and Overall Survival by Additive 5-Fluorouracil Gene Score Among TNBC Participants in the SBCSS Cohort Stratified by Whether 5-Fluorouracil was Taken

	5-Fluorouracil Taken		**5-Fluorouracil Not Taken**	
	HR (95% CI)[1]	***P***	**HR (95% CI)[1]**	***P***
		Disease-Free Survival		
Score[2]				
Group 1	Reference		Reference	
Group 2	0.87 (0.35, 2.18)	0.77	0.78 (0.22, 2.84)	0.71
Group 3	0.87 (0.34, 2.22)	0.78	0.46 (0.12, 1.78)	0.26
Group 4	0.49 (0.15, 1.59)	0.23	0.51 (0.10, 2.52)	0.40
P_{trend}	0.28		0.28	
		Overall Survival		
Score[2]				
Group 1	Reference		Reference	
Group 2	0.99 (0.42, 2.33)	0.97	1.02 (0.29, 3.54)	0.97
Group 3	0.89 (0.37, 2.15)	0.79	0.61 (0.16, 2.31)	0.46
Group 4	0.72 (0.25, 2.02)	0.53	0.61 (0.12, 3.02)	0.55
P_{trend}	0.47		0.38	

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression
[2]Score based on median cut point (higher score indicates faster metabolizer, slower activator, and poorer response)

The association between the 5-fluorouracil gene expression score and the number of cycles of 5-fluorouracil completed was evaluated as a surrogate for toxicity. For every one unit increase in the gene expression score, there was a significantly increased likelihood of completing 6 or more cycles of 5-fluorouracail (OR=1.40, 95% CI: 1.03, 1.90, p=0.03), which is

expected as those with higher scores degrade 5-fluorouracil more quickly and activate it more slowly, resulting in a longer exposure to the active metabolite.

The association between the 5-fluorouracil gene expression score and breast cancer survival among those who took 5-fluorouracil was further evaluated through additional adjustment for the number of cycles of 5-fluorouracil completed (Table 51). The results were not materially changed after adjustment for cycles completed.

Table 51: Disease-Free and Overall Survival by Additive 5-Fluorouracil Gene Score Further Adjusted for Number of Cycles of 5-Fluorouracil Completed

	Disease-Free Survival		**Overall Survival**	
	HR (95% CI)[1]	***P***	**HR (95% CI)[1]**	***P***
Score[2]				
Group 1	Reference		Reference	
Group 2	0.84 (0.34, 2.09)	0.70	0.96 (0.40, 2.26)	0.92
Group 3	0.82 (0.32, 2.09)	0.68	0.85 (0.35, 2.07)	0.72
Group 4	0.46 (0.14, 1.51)	0.20	0.69 (0.24, 1.95)	0.48
P_{trend}	0.23		0.42	

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, DUSP4 expression, and number of cycles of 5-fluorouracil
[2]Score based on median cut point (higher score indicates faster metabolizer, slower activator, and poorer response)

We stratified our results by timing of survival events, those that occurred in the first three years of follow-up and those that occurred after 3 years, among all women and restricted to those who took 5-fluorouracil (Table 52). During the first 3 years of follow-up, the 5-fluorouracil gene expression score was associated with significantly better DFS among all participants (P_{trend}=0.04); a similar pattern was observed when the analyses were restricted to only those who

took 5-fluorouracil, although the P_{trend} was not significant, possibly due to reduced sample size. These results indicate that those with a shorter exposure to the 5-fluorouracil drug had better survival. Further adjustment for the number of cycles of 5-fluorouracil completed slightly attenuated the observed association.

Table 52: Association Between Disease-Free and Overall Survival and Additive 5-Fluorouracil Gene Score Stratified by Early vs. Late Events Among All Participants and Those Who Took 5-Fluorouracil

	Events <3 years				Events ≥3 years			
	Disease-Free Survival (47 events)		Overall Survival (34 events)		Disease-Free Survival (20 events)		Overall Survival (42 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
Among All Participants Score[2]								
Group 1	Reference		Reference		Reference		Reference	
Group 2	0.88 (0.37, 2.05)	0.76	2.45 (0.71, 8.50)	0.16	0.62 (0.14, 2.82)	0.54	0.38 (0.14, 1.04)	0.06
Group 3	0.56 (0.22, 1.39)	0.21	0.75 (0.19, 3.03)	0.69	1.49 (0.39, 5.65)	0.56	0.94 (0.40, 2.20)	0.89
Group 4	0.40 (0.13, 1.22)	0.11	0.91 (0.21, 4.03)	0.91	0.88 (0.17, 4.69)	0.88	0.65 (0.22, 1.89)	0.43
P_{trend}	0.04		0.11		0.61		0.93	
Among Those Who Took 5-Fluorouracil Score[2]								
Group 1	Reference		Reference		Reference		Reference	
Group 2	1.18 (0.37, 3.70)	0.78	1.91 (0.41, 8.98)	0.41	0.31 (0.05, 1.92)	0.21	0.49 (0.15, 1.56)	0.23
Group 3	0.66 (0.19, 2.30)	0.51	0.62 (0.11, 3.58)	0.59	1.37 (0.34, 5.55)	0.66	1.07 (0.38, 3.05)	0.89
Group 4	0.56 (0.13, 2.33)	0.42	1.09 (0.19, 6.29)	0.92	0.34 (0.03, 3.53)	0.37	0.56 (0.14, 2.22)	0.41
P_{trend}	0.17		0.35		0.93		0.91	

[1]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression

[2]Score based on median cut point (higher score indicates faster metabolizer, slower activator, and poorer response)

Conclusions/Discussion

Very few studies have investigated the effect of expression of drug metabolizing enzymes on prognosis in breast cancer, in particular among TNBC patients. In this study, we evaluated the association between enzymes which metabolize the two most commonly used chemotherapy drugs in our population, cyclophosphamide and 5-fluorouracil. We further evaluated the association between these genes and breast cancer survival by creating gene expression scores. We found an inverse association between ALDH1A1 expression and DFS, although this association was observed in both those who underwent treatment with chemotherapy and those that did not, and these results have been published.[121] GSTP1 gene expression was associated with significantly better DFS for events occurring in the first three years following breast cancer diagnosis, particularly among those who took cyclophosphamide. GSTT1 gene expression was associated with significantly worse OS for events occurring in the first three years following diagnosis, particularly among those who took cyclophosphamide. The cyclophosphamide gene expression score was associated with worse OS among all TNBC patients for events occurring in the first three years after cancer diagnosis. We also found a significant association between UMPS and DFS among all TNBC patients. The association between UMPS and DFS was similar in those who took 5-fluorouracil and those who did not. The 5-fluorouracil gene expression score was significantly associated with better DFS in the first three years of follow-up.

Cyclophosphamide requires activation by CYP genes in the liver before the active metabolites can passively enter the tumor cells. Therefore, tumor tissue gene expression of these CYP genes should not play a major role in cyclophosphamide activation. In our study, we have measured the tumor tissue expression of these genes but four of these genes were expressed in

~≤10% of breast cases (CYP2B6, CYP3A5, CYP2A6, CYP2C8) and two genes were only expressed in one tumor sample each (CYP3A4 and CYP2C19). This supports the idea that these genes are not responsible for cyclophosphamide activation within the tumor cell, or that these genes do not play a key role in tumor biology in TNBC.

ALDH1A1 is involved in deactivation of the cyclophosphamide active metabolite and ALDH1A1 gene expression has been shown to be associated with ER and PR negativity and poorer breast cancer outcomes.[170-172] In our study, ALDH1A1 gene expression was positively associated with ER and PR expression (r=0.22 and r=0.30, respectively). We also found that ALDH1A1 expression was associated with better breast cancer survival in our study which we previously reported.[121]

Glutathione S-transferases are also involved in the deactivation of the cyclophosphamide active metabolite. In line with our hypothesis, GSTT1 gene expression was associated with worse OS for events occurring in the first three years following breast cancer diagnosis; a similar trend was observed for GSTM1, although the association with OS was not statistically significant. Conversely, we found an association between higher GSTP1 gene expression and better DFS. This may be a chance finding or this gene may be involved in other mechanisms which influence survival.

Due to the potentially large impact of ALDH1A1 expression on the cyclophosphamide gene expression score, we further evaluated this association by creating a score which only included the GST genes. Among all participants, we found that high GST expression (all four GST genes) was associated with significantly better DFS (highest score compared to lowest score HR=0.25, 95% CI: 0.07, 0.97, p=0.046); however, a dose-response relationship was not

observed. This score was not associated with ALDH1A1 expression and the consistence suggests that the observed association may potentially be true. When stratified by whether cyclophosphamide was taken, no significant difference was observed between the two groups. These findings suggest that GSTs may play other roles in survival among TNBC patients, although the small sample size may have limited our ability to detect an interaction.

The cyclophosphamide gene expression score was associated with an increased risk of death in the first 3 years (P_{trend}=0.03); a similar association was observed when restricted to only those who took cyclophosphamide, although the P_{trend} was no longer significant. We hypothesized that this score would be associated with worse survival since the genes included in it were associated with increased degradation of the active metabolite of cyclophosphamide and would result in a shorter exposure period. We would expect the score to be stronger in those who took the drug and survival directly following diagnosis because this is the time frame when the effects of chemotherapy are most likely to be observed, although our potential to detect a significant association may have been limited by our sample size. Among those who survived three years post-diagnosis, the association between survival and the cyclophosphamide gene expression score was in the opposite direction (better survival), although the trend was not statistically significant. The cyclophosphamide gene expression score was not associated with the number of cycles of cyclophosphamide completed and further adjustment for cycles of cyclophosphamide did not alter the observed associations.

In line with our hypotheses for genes involved in metabolism of 5-fluorouracil, TYMP expression was associated with better DFS and TYMS was associated with worse DFS among those who received 5-fluorouracil treatment, though neither reached statistical significance.

TYMS activity is essential for DNA synthesis and cell proliferation.[173] The major mechanism of action of 5-fluorouracil is to inhibit TYMS enzyme activity.[51] The TYMP enzyme catalyzes one step in the conversion of 5-fluorouracil to fluorodeoxyuridine monophosphate (FdUMP), the latter of which is crucial for inhibiting the TYMS enzyme.[51] These findings support previous studies which have shown similar associations between TYMP[130-132] and TYMS[129,133] expression and breast cancer outcomes.

We found no association between tumor-level DPYD gene expression and breast cancer survival among those who were treated with 5-fluorouracil. Although some studies have shown an association between DPYD expression and worse survival in breast cancer[128] and other cancers[174-176], other studies have found no association.[129,177,178] Most of these studies looked at protein expression, rather than gene expression; several studies have shown that DPYD protein activity and mRNA levels may not be strongly correlated, which may explain why we did not find an associaition. Alternatively, although degradation of 5-fluorouracil by DPYD occurs in all tissues, including tumor, it primarily occurs in the liver.[49,62] It may be that tumor-level expression and degradation have little effect on survival outcomes.

Our study found an association between higher UMPS expression amd worse survival. UMPS, also known as OPRT, is the main enzyme involved in the conversion of 5-fluorouracil to 5-fluorouridine monophosphate, which is essential for inhibiting TYMS, and previous studies have shown that high levels of UMPS/OPRT are associated with better survival.[179-181] However, another study found that the ratio of OPRT to DPYD, in addition to OPRT expression, may be of importance in survival in patients with metastatic CRC treated with 5-fluorouracil.[182] Among those who took 5-fluorouracil, we found no association between the ratio of UMPS/DPYD and

DFS (p=0.92); however, we did see a non-significant decrease in OS after adjustment for age at diagnosis, tumor grade, basal-like subtype, DUSP4 expression, and number of cycles of 5-fluorouracil (HR=0.47, 95% CI: 0.17, 1.33, p=0.16).

MTHFR expression was associated with significantly better DFS and OS among all participants; however, after fully adjusting the model, the association lost significance. The observed association was limited to those that did not take 5-fluorouracil as compared with those that did and the interaction was significant for OS (p=0.04). MTHFR metabolizes folate which acts as a cofactor in the inhibition of TYMS by the active metabolite of 5-fluorouracil; however, one study showed that higher gene expression levels of MTHFR were significantly associated with methylenetetrahydrofolate concentration, the form of folate which acts as a cofactor in the TS inhibition mechanism.[183] The role of MTHFR extends far beyond that of 5-fluorouracil action.[49] Folate plays a critical role in DNA synthesis and methylation.[184] Among those treated with 5-fluorouracil, there was no difference in expression of MTHFR; however, among those who were not treated with 5-fluorouracil, those who had a survival event had significantly lower levels of MTHFR. This suggests that other aspects of MTHFR activity may be the driving factor between expression and survival.

The 5-fluorouracil gene expression score was associated with significantly better DFS for the first three years following cancer diagnosis. This indicates that those with shorter exposure periods to the active metabolite of 5-fluorouracil were less likely to have a recurrence or die from breast cancer. This may be due to toxicity. When number of cycles of 5-fluorouracil completed was controlled for, the observed association was slightly attenuated and lost statistical

significance. There is a trade-off between toxicity and long-term outcome; toxicity may increase short-term mortality but may improve long-term outcome through lower rates of recurrence.

We used the PAM50 subtype predictor to classify our TNBC patients by molecular subtype based on gene expression profiling, which enabled adjustment for basal-like subtype in our analyses. However, the proportion of basal-like TNBC patients in our populations was lower than expected. It should be noted that the PAM50 score was designed to be used in samples with a global population of breast cancer patients where all 5 subtypes are present.[137] Our subtype prediction was done in our sample of Chinese TNBC patients only which may lead to misclassification of subtype or underestimation due to the inability to measure basal-like breast cancers that are not TNBC.

As previously mentioned, metabolism of chemotherapy drugs occurs outside of the breast tissue, primarily in the liver. Gene expression level in the breast tumor tissue of metabolizing genes may not be directly related to the drug exposure in target tissue. Using $\log_2$-tranformed gene expression data from from GTEx, we examined the correlation between gene expression in the breast and expression in the liver and whole blood for the genes included in this study. GSTM1 and GSTT1 expression in the breast tissue were highly significantly correlated with expression in the liver (Spearman correlation=0.82, $p<.0001$; Spearman correlation=0.65, $p<.0001$, respectively) and whole blood (Spearman correlation=0.84, $p<.0001$; Spearman correlation=0.79, $p<.0001$, respectively). This suggests that the expression level of these two genes in the breast tissue can serve as as surrogate measurements for their level in liver where most of the chemotherapy drugs are metabolized. Thus, the association we observed between these gene expressions in breast cancer tissue and worse OS for events occurring in the first three

years following breast cancer diagnosis may have more biological relevance. TYMP expression in breast tissue was also highly correlated with expression in liver tissue (Spearman correlation=0.57, $p<.0001$) and whole blood (Spearman correlation=0.18, $p=0.02$). UMPS, DPYD, TYMS, and ALDH1A1 expression in breast tissue, on the other hand, were moderately correlated with expression in the liver (Spearman correlation=0.28, $p=0.02$; Spearman correlation=0.24, $p=0.05$; Spearman correlation=0.24, $p=0.05$; Spearman correlation=0.32, $p=0.01$, respectively), but not in the blood. These findings lend some supporting evidence on the relevance of the findings of our study between tumor-level expression of these metabolizing genes and breast cancer survival outcomes. No statistically significant correlations between breast tissue gene expression of MTHFR, GSTP1, and GSTA1 and liver or blood expression were observed, which may partially explain the overall null results we found for these gene expressions in the current study. However, the level of folate, which is metabolized by MTHFR, available for the reaction inhibiting TYMS and preventing DNA replication would be on the cellular level. Therefore, the association between MTHFR and better survival among participants treated with 5-fluorouracil would not be diminished by the lack of correlation between breast and liver or blood expression of MTHFR.

The gene expression in tumor tissue in this study was measured from tumor tissue taken prior to chemotherapy initiation. Because the majority of the women included in our study had a mastectomy (95%), the influence of gene expression level in breast cancer tissue may be of less relevance if they are not correlated to those of the metablizing organ (i.e., liver). Tumor-level gene expression may be more relevant in those who undergo neoadjuvant chemotherapy treatment.

The majority of previous studies which have been published and mentioned here used protein expression whereas we evaluated gene expression. Differences in our results compared to these studies may be due to low correlation between protein and gene expression levels, which may vary by gene. Few studies have looked at tumor level gene expression, particularly in breast cancer; therefore, more studies are needed to further assess the potential clinical utility of gene expression of chemotherapy metabolizing genes.

CHAPTER VI

JOINT EFFECT BETWEEN GERMLINE VARIATION GENE SCORES AND TUMOR-LEVEL GENE EXPRESSION SCORES AND TNBC SURVIVAL

Aim 4-Specific Methods

For this aim, we used all participants with both gene expression and genotyping data available, resulting in 312 total participants; although data for all 312 participants was not available for all SNPs. We calculated Pearson's correlation coefficients and corresponding *p*-values to evaluate the association between each SNP and its corresponding gene expression as well as its corresponding gene expression score. For comparability, the correlation between the SNP and the corresponding gene expression were also evaluated in the TCGA data.

For the genes involved in cyclophosphamide metabolism, we used the SNP gene score created in Aim 2 and the gene expression score created in Aim 3. For the genes involved in 5-fluorouracil, we used the SNP gene score created in Aim 2 which only included SNPs in the DPYD, TYMP, and UMPS for comparability to the cyclophosphamide score, which only included SNPs from genes involved in activation and deactivation. Therefore, we created a new gene expression score based on median cut points that only included these same three genes. For the joint effect analysis, there were only 95 participants that had data available for both the SNP score and the gene expression score.

Participants were categorized as low or high metabolizers based on their SNP score and gene expression score, separately, and then categorized into groups: low-low, low-high, high-

low, and high-high. Those in the higher categories have longer exposure time to the drug of interest while those in the low categories have shorter exposure times.

Results

The correlation between the minor allele for each SNP and expression of it's corresponding gene as well as the total gene expression score associated with that gene among only those participants with data available for SNPs and gene expression was evaluated (Table 53). The A allele in SNP rs3764435 was moderately correlated with ALDH1A1 gene expression (r=0.20, p=0.06). The G allele in SNP rs1695 was moderately inversely associated with GSTP1 gene expression (r=-0.11, p=0.06). The C allele in the SNP rs1801159 was significantly inversely correlated with DPYD gene expression (r=-0.16, p=0.006). The G allele in the SNP rs1801265 was moderately associated with DPYD expression (r=0.11, p=0.06). The G allele in the SNP rs17376848 was significantly inversely correlated with the 5-fluorouracil gene expression score (r=-0.23, p=0.02). The G allele in the SNP rs3772809 was significantly correlated with the 5-fluorouracil gene expression score (r=0.12, p=0.03). None of the significant findings were also significant in the TCGA data; however, the direction of association was similar in all cases. SNPs in the promotor regions of the genes of interest were not included in this study. SNPs in the promotor region may play a larger role in the role of the expression levels of these genes in the tumor tissue.

There was no correlation between the cyclophosphamide gene expression score and the cyclophosphamide SNP score (r=0.06, p=0.59) or between the 5-fluorouracil gene expression score and the 5-fluorouracil SNP score (r=-0.0002, p>0.99).

Among the 95 participants with genoptyping data and gene expression score data available, there were 13 recurrences/breast cancer-specific deaths and 16 total deaths over a median follow-up of 5.3 years (range: 0.7-8.9 years). Cyclophosphamide was taken by 76 patients and 5-fluorouracil was taken by 78 patients; 75 patients took both drugs.

We evaluated the joint effect of the SNP score (Aim 2) and gene expression score (Aim 3) on DFS and OS among all participants and among only those who took the chemotherapy drug of interest for cyclophosphamide (Table 54) and 5-fluorouracil (Table 55).

Table 53: Correlations Between SNPs and Gene Expression and Gene Scores

				SBCSS Data		TCGA Data (N=887)	
				Correlation with Tumor-Level Gene Expression		Correlation with Tumor-Level Gene Expression	
Gene	**rs ID**	**N**	**MAF**	**r**	***p*[1]**	**r**	***p*[1]**
Cyclophosphamide							
Activation							
CYP2B6	rs3745274	97	T=0.19	-0.09	0.38	-0.04	0.29
CYP2C19	rs4244285	311	A=0.33	-0.05	0.36	-0.07	0.05
CYP2C19	rs4986893	311	A=0.06	-0.02	0.73	0.01	0.77
CYP2C8	rs2071426	311	C=0.07	0.01	0.84	0.001	0.98
Deactivation							
ALDH1A1	rs3764435	95	A=0.47	0.20	0.06	0.02	0.63
ALDH1A1	rs63319	95	G=0.44	-0.09	0.38	Not available	
ALDH3A1	rs2228100	n/a	C=0.42	Expression data not available		Not evaluated	
ALDH3A1	rs887241	n/a	A=0.06	Expression data not available		Not evaluated	
ALDH3A1	rs3744692	n/a	T=0.07	Expression data not available		Not evaluated	
GSTA1	rs3957357	95	A=0.14	0.03	0.76	-0.01	0.72
GSTP1	rs1695	311	G=0.20	-0.11	0.06	-0.03	0.33
Overall SNP Score	n/a	95	n/a	n/a		n/a	
5-Fluorouracil							
Activation							
TYMP	rs11479	311	A=0.21	-0.03	0.66	0.02	0.48
UMPS	rs1801019	311	C=0.18	0.05	0.39	-0.004	0.90
UMPS	rs3772809	312	G=0.06	-0.04	0.45	-0.003	0.93
Deactivation							
DPYD	rs17376848	95	G=0.10	-0.03	0.79	0.06	0.07
DPYD	rs1801159	311	C=0.27	-0.16	0.006	-0.03	0.46
DPYD	rs1801265	311	G=0.09	0.11	0.06	-0.01	0.72
DPYD	rs72728438	95	C=0.23	-0.003	0.98	-0.06	0.06
Response							
MTHFR	rs1801131	312	G=0.18	0.08	0.18	0.09	0.007
MTHFR	rs1801133	95	A=0.44	-0.06	0.59	0.01	0.66
MTHFR	rs2274976	311	T=0.09	-0.002	0.97	0.006	0.86
TYMS	rs2847153	97	A=0.36	0.06	0.55	0.06	0.07
TYMS	rs2853533	95	G=0.49	-0.04	0.70	-0.11	0.001
Overall SNP Score	n/a	95	n/a	n/a		n/a	

[1]P-values not adjusted for multiple comparisons

Table 54: Joint Effect of Cyclophosphamide Gene Expression Score and SNP Score

	Disease-Free Survival		Overall Survival	
	Gene Expression Score HR[1] (95%CI)		Gene Expression Score HR[1] (95%CI)	
SNP Gene Score	**Low Expression**	**High expression**	**Low expression**	**High expression**
Overall (n=95)				
Shorter exposure	1.00 (reference)	0.28 (0.03, 2.41)	1.00 (reference)	2.27 (0.56, 9.28)
Longer exposure	0.48 (0.14, 1.68)	0.55 (0.07, 4.59)	1.19 (0.34, 4.25)	0.91 (0.10, 8.22)
Those who Took Cyclophosphamide (n=76)				
Shorter exposure	1.00 (reference)	0.22 (0.02, 2.10)	1.00 (reference)	2.05 (0.36, 11.66)
Longer exposure	0.62 (0.16, 2.37)	0.71 (0.08, 6.17)	1.56 (0.39, 6.30)	1.31 (0.13, 12.87)

[1]Adjusted for age at diagnosis and tumor grade

Table 55: Joint Effect of 5-Fluorouracil Gene Expression Score and SNP Score

	Disease-Free Survival		Overall Survival	
	Gene Expression Score HR[1] (95%CI)		Gene Expression Score HR[1] (95%CI)	
SNP Gene Score	**Shorter exposure**	**Longer exposure**	**Shorter exposure**	**Longer exposure**
Overall (n=95)				
Shorter exposure	1.00 (reference)	0.50 (0.09, 2.71)	1.00 (reference)	0.34 (0.07, 1.67)
Longer exposure	0.98 (0.28, 3.40)	0.24 (0.03, 2.09)	0.52 (0.16, 1.73)	0.36 (0.08, 1.70)
Those who Took 5-Fluorouracil (n=78)				
Shorter exposure	1.00 (reference)	0.63 (0.11, 3.68)	1.00 (reference)	0.40 (0.08, 1.97)
Longer exposure	1.46 (0.38, 5.58)	0.24 (0.03, 2.15)	0.51 (0.13, 2.02)	0.32 (0.07, 1.55)

[1]Adjusted for age at diagnosis and tumor grade

CHAPTER VII

SUMMARY/FUTURE DIRECTIONS

Response to chemotherapy varies widely among TNBC patients and those who respond well to treatment have 5-year survival rates similar to those diagnosed with other breast cancer subtypes. Little is known about why some patients respond well while others do not; individual inherited differences in chemotherapy drug metabolism likely play a role in differences in response. A better understanding of the underlying molecular biological mechanisms is critical to address the unmet clinical need to improve the outcomes for this group of aggressive breast cancers. Ultimately, identification of determinants of chemotherapy response will guide clinical decision-making and allow for a personalized approach to chemotherapy selection for TNBC patients.

Drug metabolizing genes play an important role in the success or failure of cancer therapies.[19,185] Identification of responsive tumors may guide clinical therapeutic decisions leading to more personalized approaches for breast cancer chemotherapy.[186,187] Understanding the pharmacogenetic variables which influence prognosis could lead to more effective treatment selection and improved outcomes in breast cancer patients.[186] TNBC patients who respond well to chemotherapy and achieve a pCR have survival and recurrence rates which are similar to other breast cancer subtypes; therefore, identifying determinants of chemotherapy response has the potential to influence clinical practice and improve outcomes and TNBC patients stand to gain the most from better chemotherapy selection.

The majority of previous studies evaluating the association between genetics and chemotherapy response have evaluated one or a few polymorphisms in a given gene. It has been suggested that multigene- and pathway-oriented analysis may be a better approach,[188] although few studies have been able to evaluate these associations. We used a gene score approach to account for multiple genetic polymorphisms which influence the metabolism of a certain chemotherapy drug. This gene score approach allowed us to maximize the power of our study, particularly for less common variants, as well as to assess whether more polymorphisms in particular genes in the biological pathways of interest led to differences in survival (a dose-response relationship).

Both heritable genetic factors and tumor genomic factors may affect individual response to drug treatment.[64,189] In this study, we analyzed data on both the germline DNA as well as the gene expression profiles of the tumor tissue. This unique combination allows for a more comprehensive evaluation, both separately and jointly, of their relationship with disease-free and overall survival. This is particularly important for our study of genes involved in chemotherapy metabolism since activation and deactivation of the chemotherapy metabolites occurs both in normal tissue (typically the liver) as well as in the tumor tissue. By incorporating genetic variation in both sources, we may be able to better understand chemotherapy metabolism and cancer outcomes.

Most studies assessing the pharmacogenetics of chemotherapy drug response and cancer survival have treated breast cancer as a single disease; however, this is not the strongest approach because breast cancer is a heterogeneous disease. Discrepancies in the results of previous biomarkers studies in breast cancer may be due to differences across subtypes of breast cancer.[190]

In our study, we were able to run our gene expression analyses in TNBC patients only; however, we were unable to do so for the germline genetic variation analysis due to a limited number of TNBC patients with data available on the SNPs of interest. We plan to rerun the analyses among TNBC patients only when data is available.

Our study has several additional noteworthy strengths. The cohort that this data was collected from is a large, population-based study with comprehensive collection of information related to covariates and chemotherapy use. The validity of the data was increased through review of medical charts for disease and clinical information. Additionally, our study was done in an Asian population while the majority of previous studies have been conducted in Causasian populations. This allowed for the study of some SNPs which are extremely rare in Caucasian populations but more common among Asians (e.g. rs4986893, rs3744692, and rs3772809).

The sample size of our study limited our ability to draw conclusions for some of our findings, particularly for analyses stratified by whether the chemotherapy drug of interest was taken, and our ability to assess interactions by chemotherapy regimen. Additionally, our joint analysis of germline variation and tumor-level gene expression among TNBC patients was limited by small sample size.

In our study, we did not have information on toxicity for our patients. Patients who have longer exposure periods to the active metabolite of a chemotherapy drug may be more likely to experience an adverse reaction or toxicity event. This may occur through faster activation of the drug or slower deactivation. If the reaction is severe enough, then the chemotherapy regimen that the patient is receiving may be changed. We used number of cycles of chemotherapy as a surrogate to try to account for this in our analyses. Information on adverse reactions and toxicity

events may allow for a better understanding of individual response to chemotherapy which would affect survival outcomes. Future studies should better incorporate toxicity events into data analyses for a more comprehensive evaluation of the effect of SNPs on chemotherapy response and survival outcomes. Further, it would be interesting to incorporate data on circulating drug levels into future studies to better understand the effects of genetic and tumor variation on drug metabolism.

Extensive information on chemotherapy was available for the participants in our study. However, many of the participants in our study were taking multiple chemotherapy drugs. We looked at genes involved in the metabolism of cyclophosphamide and 5-fluorouracil but not other chemotherapy drugs. This is particularly important for those who took an anthracycline as some studies have shown anthracycline-based chemotherapy regimens to be associated with better outcomes in breast cancer patients.[15] A little over 50% of the participants in the SBCSS cohort took an anthracycline; however, further controlling for anthracycline use did not materially change our results.

A suitable validation data set for this study was not available. We attempted to validate our findings using TCGA data (see Appendix C for full write-up of TCGA findings); however, most of the findings from our study did not replicate in this population. There are several important limitations in using TCGA data as a validation set for this study. The biggest issue is that treatment information is not known for the patients included in the TCGA data set. We do not know whether the patients received chemotherapy nor do we have information on what types of chemotherapy were given. Our study focused on genes involved in chemotherapy metabolism which should only be relevant among those who took certain types of chemotherapy drugs

metabolized by the genes studied. Furthermore, the TCGA data was not a systematic collection of cases and the breast cancer cases were more likely to be diagnosed at a later stage than the women in the SBCSS cohort. Additionally, follow-up time for these patients is relatively short and the number of TNBC patients is extremely limited (n=93). The next important step for this study would be to validate our findings in a more suitable dataset in order to strengthen our results.

More studies are needed to further elucidate the genes involved in the pathways by which chemotherapy drugs are metabolized. While many of the genes involved in metabolism have been identified, these genes often play other roles in normal cell function as well as tumor progression and cancer survival. Additionally, the effect of both germline and tumor-level genetic variation may be better evaluated in a randomized trial where chemotherapy treatment is more controlled, i.e. types and dosages, combination therapy, etc. It would also be interesting to look at both tumor response and survival as outcomes to get a better understanding of the role of chemotherapy metabolizing genes on breast cancer outcomes. These types of studies may allow for genetic variation in genes involved in chemotherapy metabolism to be incorporated in clinical use to guide chemotherapy selection, but at this time our understanding is too limited.

REFERENCES

1. Pal SK, Childs BH, Pegram M. Triple negative breast cancer: unmet medical needs. *Breast cancer research and treatment.* Feb 2011;125(3):627-636.
2. Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Aug 1 2007;13(15 Pt 1):4429-4434.
3. Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review. *Histopathology.* Jan 2008;52(1):108-118.
4. Hudis CA, Gianni L. Triple-negative breast cancer: an unmet medical need. *The oncologist.* 2011;16 Suppl 1:1-11.
5. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer.* May 1 2007;109(9):1721-1728.
6. Ossovskaya V, Wang Y, Budoff A, et al. Exploring molecular pathways of triple-negative breast cancer. *Genes & cancer.* Sep 2011;2(9):870-879.
7. Dawood S. Triple-negative breast cancer: epidemiology and management options. *Drugs.* Dec 3 2010;70(17):2247-2258.
8. Bosch A, Eroles P, Zaragoza R, Vina JR, Lluch A. Triple-negative breast cancer: molecular features, pathogenesis, treatment and current lines of research. *Cancer treatment reviews.* May 2010;36(3):206-215.

9. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Aug 15 2005;11(16):5678-5685.

10. Liedtke C, Mazouni C, Hess KR, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Mar 10 2008;26(8):1275-1281.

11. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Apr 15 2007;13(8):2329-2334.

12. Huober J, von Minckwitz G, Denkert C, et al. Effect of neoadjuvant anthracycline-taxane-based chemotherapy in different biological breast cancer phenotypes: overall results from the GeparTrio study. *Breast cancer research and treatment.* Nov 2010;124(1):133-140.

13. Balko JM, Cook RS, Vaught DB, et al. Profiling of residual breast cancers after neoadjuvant chemotherapy identifies DUSP4 deficiency as a mechanism of drug resistance. *Nature medicine.* Jul 2012;18(7):1052-1059.

14. von Minckwitz G, Untch M, Blohmer JU, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* May 20 2012;30(15):1796-1804.

15. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience? *Nature reviews. Clinical oncology.* Dec 2010;7(12):683-692.

16. Bray J, Sludden J, Griffin MJ, et al. Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide. *British journal of cancer.* Mar 16 2010;102(6):1003-1009.

17. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *The New England journal of medicine.* Nov 11 2010;363(20):1938-1948.

18. Perou CM. Molecular stratification of triple-negative breast cancers. *The oncologist.* 2011;16 Suppl 1:61-70.

19. Riddick DS, Lee C, Ramji S, et al. Cancer chemotherapy and drug metabolism. *Drug metabolism and disposition: the biological fate of chemicals.* Aug 2005;33(8):1083-1096.

20. Gor PP, Su HI, Gray RJ, et al. Cyclophosphamide-metabolizing enzyme polymorphisms and survival outcomes after adjuvant chemotherapy for node-positive breast cancer: a retrospective cohort study. *Breast cancer research : BCR.* 2010;12(3):R26.

21. Wilkinson GR. Drug metabolism and variability among patients in drug response. *The New England journal of medicine.* May 26 2005;352(21):2211-2221.

22. de Jonge ME, Huitema AD, Rodenhuis S, Beijnen JH. Clinical pharmacokinetics of cyclophosphamide. *Clinical pharmacokinetics.* 2005;44(11):1135-1164.

23. Ekhart C, Doodeman VD, Rodenhuis S, Smits PH, Beijnen JH, Huitema AD. Influence of polymorphisms of drug metabolizing enzymes (CYP2B6, CYP2C9, CYP2C19, CYP3A4, CYP3A5, GSTA1, GSTP1, ALDH1A1 and ALDH3A1) on the

pharmacokinetics of cyclophosphamide and 4-hydroxycyclophosphamide. *Pharmacogenetics and genomics.* Jun 2008;18(6):515-523.

24. Whirl-Carrillo M, McDonagh EM, Hebert JM, et al. Pharmacogenomics knowledge for personalized medicine. *Clinical pharmacology and therapeutics.* Oct 2012;92(4):414-417.

25. Chang TK, Weber GF, Crespi CL, Waxman DJ. Differential activation of cyclophosphamide and ifosphamide by cytochromes P-450 2B and 3A in human liver microsomes. *Cancer research.* Dec 1 1993;53(23):5629-5637.

26. Roy P, Yu LJ, Crespi CL, Waxman DJ. Development of a substrate-activity based approach to identify the major human liver P-450 catalysts of cyclophosphamide and ifosfamide activation based on cDNA-expressed activities and liver microsomal P-450 profiles. *Drug metabolism and disposition: the biological fate of chemicals.* Jun 1999;27(6):655-666.

27. Naguib FN, el Kouni MH, Cha S. Enzymes of uracil catabolism in normal and neoplastic human tissues. *Cancer research.* Nov 1985;45(11 Pt 1):5405-5412.

28. Westbrook K, Stearns V. Pharmacogenomics of breast cancer therapy: an update. *Pharmacology & therapeutics.* Jul 2013;139(1):1-11.

29. Vianna-Jorge R, Festa-Vasconcellos JS, Goulart-Citrangulo SM, Leite MS. Functional polymorphisms in xenobiotic metabolizing enzymes and their impact on the therapy of breast cancer. *Frontiers in genetics.* 2012;3:329.

30. Ekhart C, Rodenhuis S, Smits PH, Beijnen JH, Huitema AD. An overview of the relations between polymorphisms in drug metabolising enzymes and drug transporters

and survival after cancer drug treatment. *Cancer treatment reviews.* Feb 2009;35(1):18-31.

31. Choi JY, Nowell SA, Blanco JG, Ambrosone CB. The role of genetic variability in drug metabolism pathways in breast cancer prognosis. *Pharmacogenomics.* Jun 2006;7(4):613-624.

32. Yao S, Barlow WE, Albain KS, et al. Gene polymorphisms in cyclophosphamide metabolism pathway,treatment-related toxicity, and disease-free survival in SWOG 8897 clinical trial for breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Dec 15 2010;16(24):6169-6176.

33. Rocha V, Porcher R, Fernandes JF, et al. Association of drug metabolism gene polymorphisms with toxicities, graft-versus-host disease and survival after HLA-identical sibling hematopoietic stem cell transplantation for patients with leukemia. *Leukemia.* Mar 2009;23(3):545-556.

34. Su HI, Sammel MD, Velders L, et al. Association of cyclophosphamide drug-metabolizing enzyme polymorphisms and chemotherapy-related ovarian failure in breast cancer survivors. *Fertility and sterility.* Jul 2010;94(2):645-654.

35. Petros WP, Hopkins PJ, Spruill S, et al. Associations between drug metabolism genotype, chemotherapy pharmacokinetics, and overall survival in patients with breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Sep 1 2005;23(25):6117-6125.

36. Yao S, Sucheston LE, Zhao H, et al. Germline genetic variants in ABCB1, ABCC1 and ALDH1A1, and risk of hematological and gastrointestinal toxicities in a SWOG Phase III trial S0221 for breast cancer. *The pharmacogenomics journal.* Jun 2014;14(3):241-247.

37. Ekhart C, Rodenhuis S, Smits PH, Beijnen JH, Huitema AD. Relations between polymorphisms in drug-metabolising enzymes and toxicity of chemotherapy with cyclophosphamide, thiotepa and carboplatin. *Pharmacogenetics and genomics.* Nov 2008;18(11):1009-1015.

38. Khrunin A, Ivanova F, Moisseev A, et al. Pharmacogenomics of cisplatin-based chemotherapy in ovarian cancer patients of different ethnic origins. *Pharmacogenomics.* Jan 2012;13(2):171-178.

39. Khrunin AV, Moisseev A, Gorbunova V, Limborska S. Genetic polymorphisms and the efficacy and toxicity of cisplatin-based chemotherapy in ovarian cancer patients. *The pharmacogenomics journal.* Feb 2010;10(1):54-61.

40. Zhang BL, Sun T, Zhang BN, et al. Polymorphisms of GSTP1 is associated with differences of chemotherapy response and toxicity in breast cancer. *Chinese medical journal.* Jan 2011;124(2):199-204.

41. Oliveira AL, Rodrigues FF, Santos RE, et al. GSTT1, GSTM1, and GSTP1 polymorphisms and chemotherapy response in locally advanced breast cancer. *Genetics and molecular research : GMR.* 2010;9(2):1045-1053.

42. Chacko P, Joseph T, Mathew BS, Rajan B, Pillai MR. Role of xenobiotic metabolizing gene polymorphisms in breast cancer susceptibility and treatment outcome. *Mutation research.* Mar 7 2005;581(1-2):153-163.

43. Lizard-Nacol S, Coudert B, Colosetti P, Riedinger JM, Fargeot P, Brunet-Lecomte P. Glutathione S-transferase M1 null genotype: lack of association with tumour characteristics and survival in advanced breast cancer. *Breast cancer research : BCR.* 1999;1(1):81-87.

44. Ambrosone CB, Sweeney C, Coles BF, et al. Polymorphisms in glutathione S-transferases (GSTM1 and GSTT1) and survival after treatment for breast cancer. *Cancer research.* Oct 1 2001;61(19):7130-7135.

45. Khedhaier A, Remadi S, Corbex M, et al. Glutathione S-transferases (GSTT1 and GSTM1) gene deletions in Tunisians: susceptibility and prognostic implications in breast carcinoma. *British journal of cancer.* Oct 20 2003;89(8):1502-1507.

46. Saadat M, Khalili M, Nasiri M, Rajaei M, Omidvari S, Saadat I. Clinical response to chemotherapy in locally advanced breast cancer was not associated with several polymorphisms in detoxification enzymes and DNA repair genes. *Biochemical and biophysical research communications.* Mar 2 2012;419(1):117-119.

47. Mishra A, Chandra R, Mehrotra PK, Bajpai P, Agrawal D. Glutathione S-transferase M1 and T1 polymorphism and response to neoadjuvant chemotherapy (CAF) in breast cancer patients. *Surgery today.* Apr 2011;41(4):471-476.

48. Kristensen MH, Pedersen PL, Melsen GV, Ellehauge J, Mejer J. Variants in the dihydropyrimidine dehydrogenase, methylenetetrahydrofolate reductase and thymidylate synthase genes predict early toxicity of 5-fluorouracil in colorectal cancer patients. *The Journal of international medical research.* May-Jun 2010;38(3):870-883.

49. Diasio RB, Harris BE. Clinical pharmacology of 5-fluorouracil. *Clinical pharmacokinetics.* Apr 1989;16(4):215-237.

50. Early Breast Cancer Trialists' Collaborative G. Multi-agent chemotherapy for early breast cancer. *The Cochrane database of systematic reviews.* 2002(1):CD000487.

51. Thorn CF, Marsh S, Carrillo MW, McLeod HL, Klein TE, Altman RB. PharmGKB summary: fluoropyrimidine pathways. *Pharmacogenetics and genomics.* Apr 2011;21(4):237-242.

52. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nature reviews. Cancer.* May 2003;3(5):330-338.

53. Gonzalez-Neira A. Pharmacogenetics of chemotherapy efficacy in breast cancer. *Pharmacogenomics.* Apr 2012;13(6):677-690.

54. Parker WB, Cheng YC. Metabolism and mechanism of action of 5-fluorouracil. *Pharmacology & therapeutics.* 1990;48(3):381-395.

55. Robien K, Boynton A, Ulrich CM. Pharmacogenetics of folate-related drug targets in cancer treatment. *Pharmacogenomics.* Oct 2005;6(7):673-689.

56. Jakobsen A, Nielsen JN, Gyldenkerne N, Lindeberg J. Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphism in normal tissue as predictors of fluorouracil sensitivity. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Mar 1 2005;23(7):1365-1369.

57. Schwab M, Zanger UM, Marx C, et al. Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* May 1 2008;26(13):2131-2138.

58. Cohen V, Panet-Raymond V, Sabbaghian N, Morin I, Batist G, Rozen R. Methylenetetrahydrofolate reductase polymorphism in advanced colorectal cancer: a novel genomic predictor of clinical response to fluoropyrimidine-based chemotherapy.

Clinical cancer research : an official journal of the American Association for Cancer Research. May 2003;9(5):1611-1615.

59. Heggie GD, Sommadossi JP, Cross DS, Huster WJ, Diasio RB. Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine, and bile. *Cancer research.* Apr 15 1987;47(8):2203-2206.

60. Lu Z, Zhang R, Diasio RB. Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy. *Cancer research.* Nov 15 1993;53(22):5433-5438.

61. Fleming RA, Milano G, Thyss A, et al. Correlation between dihydropyrimidine dehydrogenase activity in peripheral mononuclear cells and systemic clearance of fluorouracil in cancer patients. *Cancer research.* May 15 1992;52(10):2899-2902.

62. Maring JG, Groen HJ, Wachters FM, Uges DR, de Vries EG. Genetic factors influencing pyrimidine-antagonist chemotherapy. *The pharmacogenomics journal.* 2005;5(4):226-243.

63. Ho DH, Townsend L, Luna MA, Bodey GP. Distribution and inhibition of dihydrouracil dehydrogenase activities in human tissues using 5-fluorouracil as a substrate. *Anticancer research.* Jul-Aug 1986;6(4):781-784.

64. Tan SH, Lee SC, Goh BC, Wong J. Pharmacogenetics in breast cancer therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Dec 15 2008;14(24):8027-8041.

65. Innocenti F, Ratain MJ. Update on pharmacogenetics in cancer chemotherapy. *Eur J Cancer.* Mar 2002;38(5):639-644.

66. Seck K, Riemer S, Kates R, et al. Analysis of the DPYD gene implicated in 5-fluorouracil catabolism in a cohort of Caucasian individuals. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Aug 15 2005;11(16):5886-5892.

67. Boisdron-Celle M, Remaud G, Traore S, et al. 5-Fluorouracil-related severe toxicity: a comparison of different methods for the pretherapeutic detection of dihydropyrimidine dehydrogenase deficiency. *Cancer letters.* May 8 2007;249(2):271-282.

68. Saif MW. Dihydropyrimidine dehydrogenase gene (DPYD) polymorphism among Caucasian and non-Caucasian patients with 5-FU- and capecitabine-related toxicity using full sequencing of DPYD. *Cancer genomics & proteomics.* Mar-Apr 2013;10(2):89-92.

69. Deenen MJ, Tol J, Burylo AM, et al. Relationship between single nucleotide polymorphisms and haplotypes in DPYD and toxicity and efficacy of capecitabine in advanced colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* May 15 2011;17(10):3455-3468.

70. Cellier P, Leduc B, Martin L, et al. Phase II study of preoperative radiation plus concurrent daily tegafur-uracil (UFT) with leucovorin for locally advanced rectal cancer. *BMC cancer.* 2011;11:98.

71. Capitain O, Boisdron-Celle M, Poirier AL, Abadie-Lacourtoisie S, Morel A, Gamelin E. The influence of fluorouracil outcome parameters on tolerance and efficacy in patients with advanced colorectal cancer. *The pharmacogenomics journal.* Aug 2008;8(4):256-267.

72. Morel A, Boisdron-Celle M, Fey L, et al. Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Molecular cancer therapeutics.* Nov 2006;5(11):2895-2904.

73. Amstutz U, Farese S, Aebi S, Largiader CR. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics.* Jun 2009;10(6):931-944.

74. Terrazzino S, Cargnin S, Del Re M, Danesi R, Canonico PL, Genazzani AA. DPYD IVS14+1G>A and 2846A>T genotyping for the prediction of severe fluoropyrimidine-related toxicity: a meta-analysis. *Pharmacogenomics.* Aug 2013;14(11):1255-1272.

75. van Kuilenburg AB, Maring JG, Schalhorn A, et al. Pharmacokinetics of 5-fluorouracil in patients heterozygous for the IVS14+1G > A mutation in the dihydropyrimidine dehydrogenase gene. *Nucleosides, nucleotides & nucleic acids.* Jun 2008;27(6):692-698.

76. Kleibl Z, Fidlerova J, Kleiblova P, et al. Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. *Neoplasma.* 2009;56(4):303-316.

77. Sulzyc-Bielicka V, Binczak-Kuleta A, Pioch W, et al. 5-Fluorouracil toxicity-attributable IVS14 + 1G > A mutation of the dihydropyrimidine dehydrogenase gene in Polish colorectal cancer patients. *Pharmacological reports : PR.* Mar-Apr 2008;60(2):238-242.

78. Magnani E, Farnetti E, Nicoli D, et al. Fluoropyrimidine toxicity in patients with dihydropyrimidine dehydrogenase splice site variant: the need for further revision of dose and schedule. *Internal and emergency medicine.* Aug 2013;8(5):417-423.

79. Savva-Bordalo J, Ramalho-Carvalho J, Pinheiro M, et al. Promoter methylation and large intragenic rearrangements of DPYD are not implicated in severe toxicity to 5-fluorouracil-based chemotherapy in gastrointestinal cancer patients. *BMC cancer.* 2010;10:470.

80. Salgado J, Zabalegui N, Gil C, Monreal I, Rodriguez J, Garcia-Foncillas J. Polymorphisms in the thymidylate synthase and dihydropyrimidine dehydrogenase genes predict response and toxicity to capecitabine-raltitrexed in colorectal cancer. *Oncology reports.* Feb 2007;17(2):325-328.

81. Salgueiro N, Veiga I, Fragoso M, et al. Mutations in exon 14 of dihydropyrimidine dehydrogenase and 5-Fluorouracil toxicity in Portuguese colorectal cancer patients. *Genetics in medicine : official journal of the American College of Medical Genetics.* Mar-Apr 2004;6(2):102-107.

82. Dhawan D, Panchal H, Shukla S, Padh H. Genetic variability & chemotoxicity of 5-fluorouracil & cisplatin in head & neck cancer patients: a preliminary study. *The Indian journal of medical research.* Jan 2013;137(1):125-129.

83. Braun MS, Richman SD, Thompson L, et al. Association of molecular markers with toxicity outcomes in a randomized trial of chemotherapy for advanced colorectal cancer: the FOCUS trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Nov 20 2009;27(33):5519-5528.

84. Ceric T, Obralic N, Kapur-Pojskic L, et al. Investigation of IVS14 + 1G > A polymorphism of DPYD gene in a group of Bosnian patients treated with 5-Fluorouracil and capecitabine. *Bosnian journal of basic medical sciences / Udruzenje basicnih*

mediciniskih znanosti = Association of Basic Medical Sciences. May 2010;10(2):133-139.

85. Saif MW, Ezzeldin H, Vance K, Sellers S, Diasio RB. DPYD*2A mutation: the most common mutation associated with DPD deficiency. *Cancer chemotherapy and pharmacology.* Sep 2007;60(4):503-507.

86. Magne N, Etienne-Grimaldi MC, Cals L, et al. Dihydropyrimidine dehydrogenase activity and the IVS14+1G>A mutation in patients developing 5FU-related toxicity. *British journal of clinical pharmacology.* Aug 2007;64(2):237-240.

87. Collie-Duguid ES, Etienne MC, Milano G, McLeod HL. Known variant DPYD alleles do not explain DPD deficiency in cancer patients. *Pharmacogenetics.* Apr 2000;10(3):217-223.

88. Offer SM, Lee AM, Mattison LK, Fossum C, Wegner NJ, Diasio RB. A DPYD variant (Y186C) in individuals of african ancestry is associated with reduced DPD enzyme activity. *Clinical pharmacology and therapeutics.* Jul 2013;94(1):158-166.

89. Zaanan A, Dumont LM, Loriot MA, Taieb J, Narjoz C. A case of 5-FU-related severe toxicity associated with the p.Y186C DPYD variant. *Clinical pharmacology and therapeutics.* Feb 2014;95(2):136.

90. Lazar A, Mau-Holzmann UA, Kolb H, Reichenmiller HE, Riess O, Schomig E. Multiple organ failure due to 5-fluorouracil chemotherapy in a patient with a rare dihydropyrimidine dehydrogenase gene variant. *Onkologie.* Dec 2004;27(6):559-562.

91. Gross E, Ullrich T, Seck K, et al. Detailed analysis of five mutations in dihydropyrimidine dehydrogenase detected in cancer patients with 5-fluorouracil-related side effects. *Human mutation.* Dec 2003;22(6):498.

92. Gross E, Busse B, Riemenschneider M, et al. Strong association of a common dihydropyrimidine dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients. *PloS one.* 2008;3(12):e4003.

93. Zhang H, Li YM, Jin X. DPYD*5 gene mutation contributes to the reduced DPYD enzyme activity and chemotherapeutic toxicity of 5-FU: results from genotyping study on 75 gastric carcinoma and colon carcinoma patients. *Med Oncol.* 2007;24(2):251-258.

94. Jennings BA, Loke YK, Skinner J, et al. Evaluating predictive pharmacogenetic signatures of adverse events in colorectal cancer patients treated with fluoropyrimidines. *PloS one.* 2013;8(10):e78053.

95. Loganayagam A, Arenas Hernandez M, Corrigan A, et al. Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *British journal of cancer.* Jun 25 2013;108(12):2505-2515.

96. Fernandez-Rozadilla C, Cazier JB, Moreno V, et al. Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *The pharmacogenomics journal.* Jun 2013;13(3):209-217.

97. Zeng H, Yu H, Lu L, et al. Genetic effects and modifiers of radiotherapy and chemotherapy on survival in pancreatic cancer. *Pancreas.* Jul 2011;40(5):657-663.

98. Ruzzo A, Graziano F, Loupakis F, et al. Pharmacogenetic profiling in patients with advanced colorectal cancer treated with first-line FOLFIRI chemotherapy. *The pharmacogenomics journal.* Aug 2008;8(4):278-288.

99. Zhang Q, Zhao YP, Liao Q, et al. Associations between gene polymorphisms of thymidylate synthase with its protein expression and chemosensitivity to 5-fluorouracil in pancreatic carcinoma cells. *Chinese medical journal.* Jan 2011;124(2):262-267.

100. Morganti M, Ciantelli M, Giglioni B, et al. Relationships between promoter polymorphisms in the thymidylate synthase gene and mRNA levels in colorectal cancers. *Eur J Cancer.* Sep 2005;41(14):2176-2183.

101. Ichikawa W, Takahashi T, Suto K, Sasaki Y, Hirayama R. Orotate phosphoribosyltransferase gene polymorphism predicts toxicity in patients treated with bolus 5-fluorouracil regimen. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Jul 1 2006;12(13):3928-3934.

102. Lamas MJ, Duran G, Gomez A, et al. X-ray cross-complementing group 1 and thymidylate synthase polymorphisms might predict response to chemoradiotherapy in rectal cancer patients. *International journal of radiation oncology, biology, physics.* Jan 1 2012;82(1):138-144.

103. Paez D, Pare L, Altes A, et al. Thymidylate synthase germline polymorphisms in rectal cancer patients treated with neoadjuvant chemoradiotherapy based on 5-fluorouracil. *Journal of cancer research and clinical oncology.* Nov 2010;136(11):1681-1689.

104. Boige V, Mendiboure J, Pignon JP, et al. Pharmacogenetic assessment of toxicity and outcome in patients with metastatic colorectal cancer treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* May 20 2010;28(15):2556-2564.

105. Kawakami K, Watanabe G. Identification and functional analysis of single nucleotide polymorphism in the tandem repeat sequence of thymidylate synthase gene. *Cancer research.* Sep 15 2003;63(18):6004-6007.

106. Pullarkat ST, Stoehlmacher J, Ghaderi V, et al. Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. *The pharmacogenomics journal.* 2001;1(1):65-70.

107. Farina-Sarasqueta A, van Lijnschoten G, Rutten HJ, van den Brule AJ. Value of gene polymorphisms as markers of 5-FU therapy response in stage III colon carcinoma: a pilot study. *Cancer chemotherapy and pharmacology.* Nov 2010;66(6):1167-1171.

108. Henriquez-Hernandez LA, Murias-Rosales A, Gonzalez-Hernandez A, de Leon AC, Diaz-Chico N, Fernandez-Perez L. Distribution of TYMS, MTHFR, p53 and MDR1 gene polymorphisms in patients with breast cancer treated with neoadjuvant chemotherapy. *Cancer epidemiology.* Oct 2010;34(5):634-638.

109. Hur H, Kang J, Kim NK, et al. Thymidylate synthase gene polymorphism affects the response to preoperative 5-fluorouracil chemoradiation therapy in patients with rectal cancer. *International journal of radiation oncology, biology, physics.* Nov 1 2011;81(3):669-676.

110. Gusella M, Frigo AC, Bolzonella C, et al. Predictors of survival and toxicity in patients on adjuvant therapy with 5-fluorouracil for colorectal cancer. *British journal of cancer.* May 19 2009;100(10):1549-1557.

111. Sharp L, Little J. Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. *American journal of epidemiology.* Mar 1 2004;159(5):423-443.

112. Etienne-Grimaldi MC, Milano G, Maindrault-Goebel F, et al. Methylenetetrahydrofolate reductase (MTHFR) gene polymorphisms and FOLFOX response in colorectal cancer patients. *British journal of clinical pharmacology.* Jan 2010;69(1):58-66.

113. Lee KH, Chang HJ, Han SW, et al. Pharmacogenetic analysis of adjuvant FOLFOX for Korean patients with colon cancer. *Cancer chemotherapy and pharmacology.* Apr 2013;71(4):843-851.

114. Shrubsole MJ, Shu XO, Ruan ZX, et al. MTHFR genotypes and breast cancer survival after surgery and chemotherapy: a report from the Shanghai Breast Cancer Study. *Breast cancer research and treatment.* May 2005;91(1):73-79.

115. Pare L, Altes A, Ramon y Cajal T, et al. Influence of thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphisms on the disease-free survival of breast cancer patients receiving adjuvant 5-fluorouracil/methotrexate-based therapy. *Anti-cancer drugs.* Aug 2007;18(7):821-825.

116. Wang J, Wang X, Zhao M, et al. Potentially functional SNPs (pfSNPs) as novel genomic predictors of 5-FU response in metastatic colorectal cancer patients. *PloS one.* 2014;9(11):e111694.

117. Chang JC, Makris A, Gutierrez MC, et al. Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. *Breast cancer research and treatment.* Mar 2008;108(2):233-240.

118. Kolacinska A, Fendler W, Szemraj J, et al. Gene expression and pathologic response to neoadjuvant chemotherapy in breast cancer. *Molecular Biology Reports.* July 2012;39(7):7435-7441.

119. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Oct 10 2005;23(29):7265-7277.

120. Khoury T, Ademuyiwa FO, Chandraseekhar R, et al. Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to neoadjuvant chemotherapy. *Modern Pathology.* March 2012;25(3):388-397.

121. Liu Y, Baglia M, Zheng Y, et al. ALDH1A1 mRNA expression in association with prognosis of triple-negative breast cancer. *Oncotarget.* Oct 8 2015.

122. Nogami T, Shien T, Tanaka T, et al. Expression of ALDH1 in axillary lymph node metastases is a prognostic factor of poor clinical outcome in breast cancer patients with 1-3 lymph node metastases. *Breast cancer.* Jan 2014;21(1):58-65.

123. Tiezzi DG, Clagnan WS, Mandarano LR, et al. Expression of aldehyde dehydrogenase after neoadjuvant chemotherapy is associated with expression of hypoxia-inducible factors 1 and 2 alpha and predicts prognosis in locally advanced breast cancer. *Clinics.* May 2013;68(5):592-598.

124. Zhou L, Luo Y, Li K, et al. Molecular markers of therapeutic resistance in breast cancer. *Human Pathology.* July 2013;44(7):1421-1428.

125. Miyake T, Nakayama T, Naoi Y, et al. GSTP1 expression predicts poor pathological complete response to neoadjuvant chemotherapy in ER-negative breast cancer. *Cancer science.* May 2012;103(5):913-920.

126. Peters WH, Roelofs HM, van Putten WL, Jansen JB, Klijn JG, Foekens JA. Response to adjuvant chemotherapy in primary breast cancer: no correlation with expression of glutathione S-transferases. *Br J Cancer.* Jul 1993;68(1):86-92.

127. Arun BK, Granville LA, Yin G, et al. Glutathione-s-transferase-pi expression in early breast cancer: association with outcome and response to chemotherapy. *Cancer Invest.* Jun 2010;28(5):554-559.

128. Horiguchi J, Takei H, Koibuchi Y, et al. Prognostic significance of dihydropyrimidine dehydrogenase expression in breast cancer. *British Journal of Cancer.* 21 Jan 2002;86(2):222-225.

129. Yu Z, Sun J, Zhen J, Zhang Q, Yang Q. Thymidylate synthase predicts for clinical outcome in invasive breast cancer. *Histology and Histopathology.* July 2005;20(3):871-878.

130. Fox SB, Engels K, Comley M, et al. Relationship of elevated tumour thymidine phosphorylase in node-positive breast carcinomas to the effects of adjuvant CMF. *Annals of Oncology.* March 1997;8(3):271-275.

131. Tominaga T, Toi M, Ohashi Y, Abe O, Group -BCS. Prognostic and predictive value of thymidine phosphorylase activity in early-stage breast cancer patients. *Clin Breast Cancer.* Apr 2002;3(1):55-64.

132. Yang Q, Barbareschi M, Mori I, et al. Prognostic value of thymidine phosphorylase expression in breast carcinoma. *International journal of cancer. Journal international du cancer.* Feb 1 2002;97(4):512-517.

133. Aki F, Bando Y, Takahashi T, et al. A retrospective study on TS mRNA expression and prediction of the effects of adjuvant oral 5-fluorouracil in breastcancer. *Oncology Letters.* November-December 2010;1(6):981-987.

134. Kolacinska A, Fendler W, Szemraj J, et al. Gene expression and pathologic response to neoadjuvant chemotherapy in breast cancer. *Molecular biology reports.* Jul 2012;39(7):7435-7441.

135. Khoury T, Ademuyiwa FO, Chandrasekhar R, et al. Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to

neoadjuvant chemotherapy. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* Mar 2012;25(3):388-397.

136. Zhou L, Luo Y, Li K, et al. Molecular markers of therapeutic resistance in breast cancer. *Human pathology.* Jul 2013;44(7):1421-1428.

137. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Mar 10 2009;27(8):1160-1167.

138. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America.* Sep 11 2001;98(19):10869-10874.

139. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine.* Dec 19 2002;347(25):1999-2009.

140. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* Jan 31 2002;415(6871):530-536.

141. Caan BJ, Sweeney C, Habel LA, et al. Intrinsic subtypes from the PAM50 gene expression assay in a population-based breast cancer survivor cohort: prognostication of short- and long-term outcomes. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* May 2014;23(5):725-734.

142. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen

receptor-positive breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Nov 1 2010;16(21):5222-5232.

143. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *The American journal of pathology.* Dec 2002;161(6):1991-1996.

144. de Ruijter TC, Veeck J, de Hoon JP, van Engeland M, Tjan-Heijnen VC. Characteristics of triple-negative breast cancer. *Journal of cancer research and clinical oncology.* Feb 2011;137(2):183-192.

145. Bertucci F, Finetti P, Cervera N, et al. How basal are triple-negative breast cancers? *International journal of cancer. Journal international du cancer.* Jul 1 2008;123(1):236-240.

146. Morris SR, Carey LA. Gene expression profiling in breast cancer. *Current opinion in oncology.* Nov 2007;19(6):547-551.

147. Rakha EA, Elsheikh SE, Aleskandarany MA, et al. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Apr 1 2009;15(7):2302-2310.

148. Bertucci F, Finetti P, Viens P, Birnbaum D. Difference in therapeutic response between basal and nonbasal triple-negative breast cancers. *The oncologist.* 2013;18(9):1060-1061.

149. Shu XO, Zheng Y, Cai H, et al. Soy food intake and breast cancer survival. *JAMA.* Dec 9 2009;302(22):2437-2443.

150. Shu XO, Gao YT, Cai Q, et al. Genetic polymorphisms in the TGF-beta 1 gene and breast cancer survival: a report from the Shanghai Breast Cancer Study. *Cancer research.* Feb 1 2004;64(3):836-839.

151. Yang G, Shu XO, Ruan ZX, et al. Genetic polymorphisms in glutathione-S-transferase genes (GSTM1, GSTT1, GSTP1) and survival after chemotherapy for invasive breast carcinoma. *Cancer.* Jan 1 2005;103(1):52-58.

152. Fu Z, Deming SL, Fair AM, et al. Well-done meat intake and meat-derived mutagen exposures in relation to breast cancer risk: the Nashville Breast Health Study. *Breast cancer research and treatment.* Oct 2011;129(3):919-928.

153. Baglia ML, Cai Q, Zheng Y, et al. Dual specificity phosphatase 4 gene expression in association with triple-negative breast cancer outcome. *Breast cancer research and treatment.* Nov 2014;148(1):211-220.

154. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature biotechnology.* Mar 2008;26(3):317-325.

155. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome research.* Jun 2002;12(6):996-1006.

156. Exome Aggregation Browser (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org).

157. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. *Nucleic acids research.* Jan 2015;43(Database issue):D662-669.

158. Cai Q, Zhang B, Sung H, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nature genetics.* Aug 2014;46(8):886-890.

159. Zhang Y, Long J, Lu W, et al. Rare coding variants and breast cancer risk: evaluation of susceptibility Loci identified in genome-wide association studies. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* Apr 2014;23(4):622-628.

160. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics.* Aug 2012;44(8):955-959.

161. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics.* Mar 1 2015;31(5):782-784.

162. Buchholz TA, Stivers DN, Stec J, et al. Global gene expression changes during neoadjuvant chemotherapy for human breast cancer. *Cancer journal.* Nov-Dec 2002;8(6):461-468.

163. Hannemann J, Oosterkamp HM, Bosch CA, et al. Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* May 20 2005;23(15):3331-3342.

164. American_Cancer_Society. A Guide to Chemotherapy. 2014.

165. Hosmer D, Lemeshow S, and May S. *Applied Survival Analysis: Regression Modeling of Time to Event Data.* 2nd ed: Wiley-Interscience; 2008.

166. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *Journal of clinical*

oncology : official journal of the American Society of Clinical Oncology. Feb 15 2001;19(4):980-991.

167. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *The New England journal of medicine.* Jun 26 2008;358(26):2796-2803.

168. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* Nov 1 2012;491(7422):56-65.

169. Baglia ML, Cai Q, Zheng Y, et al. Dual specificity phosphatase 4 gene expression in association with triple-negative breast cancer outcome. *Breast Cancer Research and Treatment.* 14 Oct 2014;148(1):211-220.

170. Ginestier C, Hur MH, Charafe-Jauffret E, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell stem cell.* Nov 2007;1(5):555-567.

171. Nalwoga H, Arnes JB, Wabinga H, Akslen LA. Expression of aldehyde dehydrogenase 1 (ALDH1) is associated with basal-like markers and features of aggressive tumours in African breast cancer. *British journal of cancer.* Jan 19 2010;102(2):369-375.

172. Morimoto K, Kim SJ, Tanei T, et al. Stem cell marker aldehyde dehydrogenase 1-positive breast cancers are characterized by negative estrogen receptor, positive human epidermal growth factor receptor type 2, and high Ki67 expression. *Cancer science.* Jun 2009;100(6):1062-1068.

173. Radparvar S, Houghton PJ, Houghton JA. Characteristics of thymidylate synthase purified from a human colon adenocarcinoma. *Archives of biochemistry and biophysics.* Jan 1988;260(1):342-350.

174. Salonga D, Danenberg KD, Johnson M, et al. Colorectal tumors responding to 5-fluorouracil have low gene expression levels of dihydropyrimidine dehydrogenase, thymidylate synthase, and thymidine phosphorylase. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Apr 2000;6(4):1322-1327.

175. Goto T, Shinmura K, Yokomizo K, et al. Expression levels of thymidylate synthase, dihydropyrimidine dehydrogenase, and thymidine phosphorylase in patients with colorectal cancer. *Anticancer Res.* May 2012;32(5):1757-1762.

176. Ciaparrone M, Quirino M, Schinzari G, et al. Predictive role of thymidylate synthase, dihydropyrimidine dehydrogenase and thymidine phosphorylase expression in colorectal cancer patients receiving adjuvant 5-fluorouracil. *Oncology.* 2006;70(5):366-377.

177. Ikeguchi M, Makino M, Kaibara N. Thymidine phosphorylase and dihydropyrimidine dehydrogenase activity in colorectal carcinoma and patients prognosis. *Langenbeck's archives of surgery / Deutsche Gesellschaft fur Chirurgie.* Oct 2002;387(5-6):240-245.

178. Kornmann M, Link KH, Galuba I, et al. Association of time to recurrence with thymidylate synthase and dihydropyrimidine dehydrogenase mRNA expression in stage II and III colorectal cancer. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract.* May-Jun 2002;6(3):331-337.

179. Ochiai T, Nishimura K, Noguchi H, et al. Prognostic impact of orotate phosphoribosyl transferase activity in resectable colorectal cancers treated by 5-fluorouracil-based adjuvant chemotherapy. *Journal of surgical oncology.* Jul 1 2006;94(1):45-50.

180. Hahnvajanawong C, Chaiyagool J, Seubwai W, et al. Orotate phosphoribosyl transferase mRNA expression and the response of cholangiocarcinoma to 5-fluorouracil. *World journal of gastroenterology.* Aug 14 2012;18(30):3955-3961.

181. Yasumatsu R, Nakashima T, Komune S. Overexpression of the orotate phosphoribosyl-transferase gene enhances the effect of 5-Fluorouracil in head and neck squamous cell carcinoma in vitro. *Journal of oncology.* 2012;2012:649605.

182. Ichikawa W, Uetake H, Shirota Y, et al. Both gene expression for orotate phosphoribosyltransferase and its ratio to dihydropyrimidine dehydrogenase influence outcome following fluoropyrimidine-based chemotherapy for metastatic colorectal cancer. *Br J Cancer.* Oct 20 2003;89(8):1486-1492.

183. Odin E, Wettergren Y, Carlsson G, et al. Expression and clinical significance of methylenetetrahydrofolate reductase in patients with colorectal cancer. *Clinical colorectal cancer.* Jan 2006;5(5):344-349.

184. Calvert H. An overview of folate metabolism: features relevant to the action and toxicities of antifolate anticancer agents. *Seminars in oncology.* Apr 1999;26(2 Suppl 6):3-10.

185. Shastry BS. Pharmacogenetics and the concept of individualized medicine. *The pharmacogenomics journal.* Jan-Feb 2006;6(1):16-21.

186. Stearns V, Davidson NE, Flockhart DA. Pharmacogenetics in the treatment of breast cancer. *The pharmacogenomics journal.* 2004;4(3):143-153.

187. Akhdar H LC, Aninat C, and More, F. Anticancer Drug Metabolism: Chemotherapy Resistance and New Therapeutic Approaches, Topics on Drug Metabolism. In: Paxton J, ed: InTech; 2012.

188. Afzal S, Gusella M, Vainer B, et al. Combinations of polymorphisms in genes involved in the 5-Fluorouracil metabolism pathway are associated with gastrointestinal toxicity in chemotherapy-treated colorectal cancer patients. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Jun 1 2011;17(11):3822-3829.

189. Dowsett M, Dunbier AK. Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Dec 15 2008;14(24):8019-8026.

190. Fuksa L, Micuda S, Grim J, Ryska A, Hornychova H. Predictive biomarkers in breast cancer: their value in neoadjuvant chemotherapy. *Cancer investigation.* Nov 2012;30(9):663-678.

APPENDIX

A. Methods for Systematic Review

Based on the PRISMA guidelines, a systematic review was conducted to identify studies investigating associations between expression of chemotherapy metabolizing genes and breast cancer outcomes. The following databases were searched: PubMed and EMBASE. The main PubMed search was: ("breast neoplasms"[MeSH Major Topic] OR "breast cancer"[All Fields]) AND "drug therapy"[Subheading] AND expression[tiab] AND (survival[tiab] OR mortality[tiab] OR disease-free[tiab] OR recurrence[tiab] OR relapse[tiab] OR prognosis[tiab] OR death[tiab]). Limitations included human, female, and English language. Additionally, ISI Web of Science was used to do a cited reference search and the reference lists of relevant studies were searched to identify additional studies.

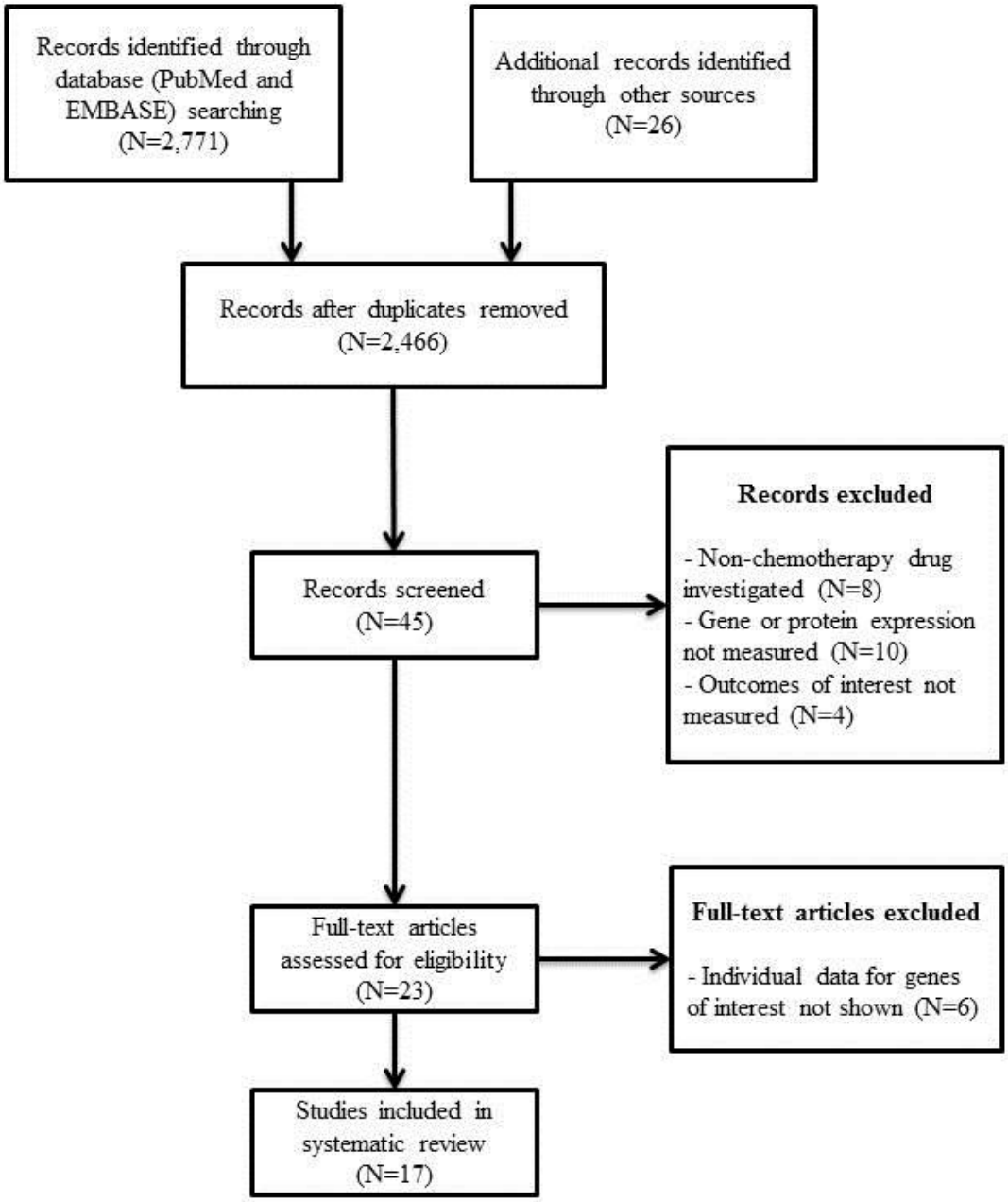
Observational studies or randomized trials with follow-up were included in the present review. Inclusion criteria included tumor-level protein or gene expression measurement of at least one chemotherapy metabolizing gene (cytochrome P450s (CYP), glutathione-S-transferases (GSTs), aldehyde dehydrogenases (ALDHs) which play a role in the metabolism of the majority of chemotherapy drugs or DPYD, MTHFR, or TYMP which metabolize fluoropyrimidines). Studies which assessed overall survival, disease-free or progression-free survival, or time to recurrence were included. Additionally, we also included studies which looked at pathologic complete response, response rate, or clinical response as outcomes.

Using Endnote software, duplicate studies selected from more than one database were removed. The study titles were screened and abstracts were reviewed for inclusion in the present

review. Full manuscripts of all identified studies were reviewed for inclusion. The data was extracted from the selected studies into Excel spreadsheets. Data extracted for this review included: author, title of article, citation, publication type, country of origin, study design, study inclusion/exclusion criteria, number of participants, breast cancer characteristics (stage, subtype information), expression measurement method, outcome measurement, definition of outcome measurement, length of follow-up, loss to follow-up, type of analysis, and results of analysis.

The PRISMA flow diagram is shown in Figure A1. A total of 2,466 citations were identified through PubMed, Web of Science, and reference list searches. 2,423 citations were excluded by title due to chemotherapy genes of interest not included in the study, study done in non-breast cancer population, or outcomes of interest were not measured. Of the 45 abstracts reviewed, 22 were excluded due to drug of interest was not chemotherapy agent (n=8), gene or protein expression was not measured (n=10), or outcomes of interest were not measured (n=4) and the remaining 23 full-text articles were obtained. After full text review, 17 studies (16 observational studies and 1 randomized trial) met inclusion criteria and were included in the systematic review.

Figure A1: PRISMA Flow Diagram of Study Selection Process



B. Quality Control Checks of Allele Frequencies by Study

Legend for Table B1

Data not Available
Overlapping Data: Imputed data only used when genotype data missing
Overlapping Data: Imputed data not needed
Imputed data had poor imputation quality (R^2<0.3)

Table B1: Allele Frequencies by Study

Gene	rs ID	Reference Allele[1]	Alternate Allele[1]	Alternate Allele Frequency in Asians[1,2]	Alternate Allele Frequency from ExomeChip (genotyped)	Alternate Allele Frequency from ExomeSeq (genotyped)	Alternate Allele Frequency from GWAS (genotyped)	Alternate Allele Frequency from ExomeChip (imputed)	Alternate Allele Frequency from GWAS (imputed)
Cyclophosphamide									
Activation									
CYP2B6	rs3745274	G	T	T=0.18		0.22	0.19		0.19
CYP2C19	rs4244285	G	A	A=0.33		0.36		0.33	0.33
CYP2C19	rs4986893	G	A	A=0.05	0.05	0.06			0.06
CYP2C8	rs2071426	T	C	C=0.07	0.07				0.06
Deactivation									
ALDH1A1	rs3764435	A	C	C=0.54					0.54
ALDH1A1	rs63319	G	T	T=0.58					0.56
ALDH3A1	rs2228100	G	C	C=0.43	0.42	0.42	0.42	0.42	
ALDH3A1	rs3744692	C	T	T=0.06	0.07	0.07		0.07	0.07
ALDH3A1	rs887241	A	C	C=0.94	0.94	0.93	0.94		
GSTA1	rs3957357	A	G	G=0.88					0.86
GSTP1	rs1695	A	G	G=0.17	0.19	0.20			0.20
5-Fluorouracil									
Activation									
TYMP	rs11479	G	A	A=0.25	0.22			0.22	0.20
UMPS	rs1801019	G	C	C=0.19	0.18	0.23			0.18
UMPS	rs3772809	A	G	G=0.06	0.06	0.05	0.06	0.06	0.06
Deactivation									
DPYD	rs17376848	A	G	G=0.12		0.19			0.10
DPYD	rs1801159	T	C	C=0.26	0.27	0.26			0.26
DPYD	rs1801265	G	A	A=0.94	0.92	0.92			0.90
DPYD	rs72728438	T	C	C=0.20					0.23
Other									
MTHFR	rs1801131	T	G	G=0.19		0.23	0.18	0.18	0.18
MTHFR	rs1801133	G	A	A=0.37		0.38			0.44
MTHFR	rs2274976	C	T	T=0.10	0.09	0.11			0.09
TYMS	rs2847153	G	A	A=0.40			0.36		
TYMS	rs2853533	G	C	C=0.54		0.52			0.51

[1]Source: HaploReg v3, forward strand alleles
[2]Source: 1000 Genomes

C. Results from Validation in Breast Cancer Cases from The Cancer Genome Atlas (TCGA)

There were 887 TCGA breast cancer cases available with gene expression data available and 87 deaths. There were 887 available with genotyping data for the SNPs investigated in this study and 93 deaths. Note that the populations for these differ by 39 individuals. The average follow-up time for the TCGA breast cancer cases was 1.2 years (range: 0.0-18.6 years). Available demographic and clinical predictors for breast cancer survival in TCGA breast cancer cases were tabulated for those included in the genotyping analyses and gene expression analyses (Table C1).

Table C1:Demographic and Clinical Predictors for Breast Cancer Survival in Breast Cancer Cases from The Cancer Genome Atlas (TCGA)

	Gene Expression Data	Genotyping Data
	N=887	N=887
	mean(SD) or N(%)	
Age at Diagnosis	58.3 (13.1)	58.0 (13.1)
Survival Time (years)	2.4 (2.9)	2.4 (2.9)
Deaths	87	93
Stage		
I	153 (17.3)	151 (17.0)
II	507 (57.2)	506 (57.1)
III	198 (22.3)	200 (22.6)
IV	15 (1.7)	15 (1.7)
Unknown	14 (1.6)	15 (1.7)
ER Status		
Positive	646 (72.8)	648 (73.1)
Negative	197 (22.2)	197 (22.2)
Unknown	44 (5.0)	42 (4.7)
PR Status		
Positive	561 (63.3)	563 (63.5)
Negative	279 (31.5)	279 (31.5)
Unknown	47 (5.3)	45 (5.1)
HER2 Status		
Positive	133 (15.0)	135 (15.2)
Borderline	134 (15.1)	136 (15.3)
Negative	461 (52.0)	470 (53.0)
Unknown	159 (17.9)	146 (16.5)

The allele frequencies for the SBCSS and TCGA breast cancer cases for the SNPs included in this study were compared to the established allele frequencies from the 1000 Genomes project (Table C2). All SNPs were similar in frequency to the relevant populations.

Table C2: Frequency of Alleles in SBCSS and TCGA Breast Cancer Cases

Gene	rs ID	Alleles[1,2]	Allele Frequency in Asians[1]	Allele Frequency in SBCSS Study Population	Allele Frequency in European[1]	Frequency in TCGA Study Population
Cyclophosphamide						
Activation						
CYP2B6	rs3745274	G, T	G=0.82	G=0.81	G=0.77	G=0.74
CYP2C19	rs4244285	G, A	A=0.33	A=0.33	A=0.15	A=0.16
CYP2C19	rs4986893	G, A	G=0.95	G=0.94	G=1.00	G=0.993
CYP2C8	rs2071426	T, C	T=0.93	T=0.93	T=0.72	T=0.71
Deactivation						
ALDH1A1	rs3764435	A, C	C=0.54	C=0.53	C=0.47	C=0.45
ALDH1A1	rs63319	G, T	G=0.42	G=0.44	G=0.49	Not Available
ALDH3A1	rs2228100	G, C	G=0.57	G=0.58	G=0.77	G=0.70
ALDH3A1	rs887241	A, C	C=0.94	C=0.94	C=0.64	C=0.68
ALDH3A1	rs3744692	C, T	T=0.06	T=0.07	T=0.00	T=0.002
GSTA1	rs3957357	A, G	G=0.88	G=0.86	G=0.60	G=0.62
GSTP1	rs1695	A, G	A=0.83	A=0.80	A=0.68	A=0.66
5-Fluorouracil						
Activation						
TYMP	rs11479	G, A	A=0.25	A=0.21	A=0.08	A=0.11
UMPS	rs1801019	G, C	C=0.19	C=0.18	C=0.14	C=0.17
UMPS	rs3772809	A, G	A=0.94	A=0.94	A=0.99	A=0.99
Deactivation						
DPYD	rs17376848	A, G	G=0.12	G=0.10	G=0.04	G=0.05
DPYD	rs1801159	T, C	T=0.74	T=0.73	T=0.83	T=0.81
DPYD	rs1801265	G, A	A=0.94	A=0.91	A=0.78	A=0.75
DPYD	rs72728438	T, C	T=0.80	T=0.77	T=0.81	T=0.81
Other						
MTHFR	rs1801131	T, G	G=0.19	G=0.18	G=0.32	G=0.28
MTHFR	rs1801133	G, A	G=0.63	G=0.56	G=0.65	G=0.69
MTHFR	rs2274976	C, T	C=0.90	C=0.91	C=0.96	C=0.95
TYMS	rs2847153	G, A	A=0.40	A=0.36	A=0.20	A=0.22
TYMS	rs2853533	G, C	G=0.46	G=0.49	G=0.86	G=0.80

[1]Source: 1000 Genomes
[2]Forward strand alleles

Cyclophosphamide

The association between each SNP in genes involved in metabolism of cyclophosphamide and OS among all breast cancer cases in the TCGA were evaluated (Table C3). The alleles were coded in the same way as for the SBCSS data (even when the minor allele differed) for ease of comparison of the results. After adjustment for age at diagnosis, TNM stage, ER, PR, and HER2 status, the C allele in the SNP rs887241 in the ALDH3A1 gene was marginally associated with poorer overall survival (HR=1.33, 95% CI: 0.99, 1.79, p=0.06) and the C allele in the SNP rs3764435 in the ALDH1A1 gene was marginally associated with poorer overall survival (HR=0.77, 95% CI: 0.56, 1.05, p=0.10). The minor allele frequency for the rs4986893 and rs3744692 SNPs were too low to be evaluated (MAF = 0.007 and 0.002, respectively). Data was not available for the SNP rs63319 in the ALDH1A1 gene. No other associations between SNPs involved in cyclophosphamide metabolism and overall survival among TCGA breast cancer cases were observed. Neither of the observed marginal associations was observed in the SBCSS data.

Each of the three gene scores created for SNPs in genes involved in cyclophosphamide metabolism were evaluated for their association with DFS and OS (Table C4). None of the gene scores were associated with overall survival among TCGA breast cancer cases.

When the results were stratified by the timing of survival events, those that occurred in the first three years of follow-up and those that occurred after three years, the activation score and total score were in the expected direction (HR<1) for among those who had events in the first three years but not in those who had later events, although the interaction was not significant

(Table C5). The results for those who had earlier events in the TCGA study population were very similar to those who had earlier events in the SBCSS study population.

Table C3: Associations Between Individual SNPs in Cyclophosphamide Metabolizing Genes and Overall Survival in TCGA Breast Cancer Cases

		TCGA				SBCSS			
Gene	**rs ID**	**Allele Frequency**	**N**	**HR (95% CI)[1]**	***P***	**Allele Frequency**	**N**	**HR (95% CI)[2]**	***P***
Activation									
CYP2B6	rs3745274	G=0.74	885	1.19 (0.84, 1.68)	0.33	G=0.81	1144	1.10 (0.79, 1.55)	0.57
CYP2C19	rs4244285	A=0.16	885	0.79 (0.51, 1.23)	0.29	A=0.33	3736	0.94 (0.82, 1.08)	0.39
CYP2C19	rs4986893	G=0.993	885	MAF too low		G=0.94	3736	0.75 (0.58, 0.96)	0.02
CYP2C8	rs2071426	T=0.71	885	1.16 (0.84, 1.61)	0.37	T=0.93	3736	0.79 (0.62, 1.01)	0.06
Deactivation									
ALDH1A1	rs3764435	C=0.45	885	0.77 (0.56, 1.05)	0.10	C=0.53	1124	1.18 (0.89, 1.57)	0.25
ALDH1A1	rs63319			Not Available		G=0.44	1124	1.05 (0.78, 1.40)	0.74
ALDH3A1	rs2228100	G=0.70	885	0.84 (0.61, 1.16)	0.30	G=0.48	3739	1.04 (0.91, 1.19)	0.57
ALDH3A1	rs887241	C=0.68	885	1.33 (0.99, 1.79)	0.06	C=0.94	3739	1.09 (0.82, 1.46)	0.54
ALDH3A1	rs3744692	T=0.002	885	MAF too low		T=0.07	3736	1.20 (0.95, 1.51)	0.13
GSTA1	rs3957357	G=0.62	885	1.11 (0.82, 1.50)	0.51	G=0.86	1124	1.02 (0.70, 1.49)	0.92
GSTP1	rs1695	A=0.66	885	1.15 (0.80, 1.64)	0.45	A=0.80	3736	1.07 (0.90, 1.26)	0.44

[1]Adjusted for age at diagnosis, tumor grade, ER, PR, and HER2 status

[2]Adjusted for age at diagnosis, tumor grade, ER, PR, and HER2 status

Table C4: Association Between Cyclophosphamide Gene Scores and Overall Survival in TCGA Breast Cancer Cases

Role in Metabolism	# of SNPs	TCGA N	TCGA HR (95% CI)[3]	TCGA *P*	SBCSS N	SBCSS HR (95% CI)[4]	SBCSS *P*
Activation[1]	4	885	1.06 (0.88, 1.29)	0.54	1141	0.94 (0.78, 1.12)	0.47
Deactivation[2]	6	885	1.03 (0.89, 1.18)	0.70	1124	1.03 (0.92, 1.15)	0.10
Total Score[1,2]	10	885	1.00 (0.89, 1.12)	0.96	1124	0.96 (0.88, 1.06)	0.44

[1]Includes SNPs from CYP genes

[2]Includes SNPs from ALDH and GST genes

[3]Adjusted for age at diagnosis, TNM stage, and ER, PR, and HER2 status

[4]Adjusted for age at diagnosis, tumor grade, and ER, PR, and HER2 status

Table C5: Association Between Cyclophosphamide Gene Scores and Overall Survival Stratified by Early vs Late Events in TCGA Breast Cancer Cases

		TCGA				SBCSS	
		Event <3 years		Event ≥3 years		Event <3 years	Event ≥3 years
Role in Metabolism	# of SNPs	Events/ Total N	HR (95% CI)[3]	Events/ Total N	HR (95% CI)[3]	HR (95% CI)[3]	HR (95% CI)[3]
Activation[1]	4	34/885	0.82 (0.60, 1.11)	59/277	1.19 (0.92, 1.54)	1.03 (0.76, 1.40)	0.88 (0.71, 1.10)
Deactivation[2]	6	34/885	1.04 (0.84, 1.28)	59/277	1.05 (0.87, 1.26)	1.06 (0.89, 1.28)	1.01 (0.88, 1.16)
Total Score[1,2]	10	34/885	0.91 (0.76, 1.09)	59/277	1.04 (0.89, 1.21)	0.94 (0.84, 1.05)	0.96 (0.86, 1.08)

[1]Includes SNPs from CYP genes

[2]Includes SNPs from ALDH and GST genes

[3]Adjusted for age at diagnosis, TNM stage, ER, PR, and HER2 status

The gene expression levels of each of the genes involved in cyclophosphamide metabolism among TCGA breast cancer cases are shown in Table C6. CYP3A4, CYP2C9, and CYP2C19 were expressed in a very low percentage of TCGA tumor tissue samples, similar to what was observed in the SBCSS data. However, unlike the SBCSS tumor tissue, CYP2B6, CYP3A5, CYP2A6, and CYP2C8 were expressed in more than half of the tumor samples. This may be due to differences in breast cancer subtype of the populations or timing of chemotherapy treatment, the latter of which is unknown for the TCGA population.

ALDH1A1 and GSTP1were expressed in all tumor tissue samples from TCGA breast cancer cases. GSTA1, GSTT1, and GSTM1 were expressed in a smaller percentage of tumors, though these percentages were higher than observed in the SBCSS population (SBCSS: GSTA1 10.0%, GSTT1 44.0%, and GSTM1 37.6%). GST deletions are more common in Asian populations than European populations so this was expected.

Table C6: Log-Transformed Expression Levels of Cyclophosphamide Metabolizing Genes in TCGA Breast Cancer Cases

Gene	% Expressed	Mean	SD	Median	Minimum	Maximum
CYP2B6	52.5%	2.14	1.88	1.84	0	9.88
CYP3A4	11.0%	0.26	0.59	0	0	7.52
CYP2C9	11.1%	0.42	0.93	0	0	7.22
CYP3A5	92.9%	2.39	1.35	2.32	0	8.88
CYP2A6	62.0%	3.33	3.75	2.15	0	18.24
CYP2C8	74.7%	2.48	1.98	2.30	0	10.93
CYP2C19	1.9%	0.25	0.65	0	0	3.29
ALDH1A1	100.0%	8.32	1.38	8.37	2.99	12.88
GSTP1	100.0%	10.96	1.56	10.82	6.35	15.59
GSTA1	61.2%	2.31	2.88	1.07	0	13.80
GSTT1	83.0%	7.31	3.76	9.03	0	11.92
GSTM1	81.7%	5.38	4.59	5.65	0	15.39

The gene expression levels for genes involved in metabolism of cyclophosphamide and overall survival among all TCGA breast cancer cases and TCGA TNBC cases only were evaluated (Table C7). No significant associations were observed between any of the genes and survival in the TCGA, either among all breast cancer cases or TNBC only.

No association was observed between the additive cyclophosphamide gene score and OS, either among all breast cancer cases or TNBC only (Table C8).

When stratified by events in the first three years and events occurring after three years, no clear differences were observed among all breast cancer cases and individual gene expression levels or the combined gene expression score (Table C9).

Table C7: Overall Survival by Expression of Genes Involved in Cyclophosphamide Metabolism in TCGA Breast Cancer Cases

	TCGA					SBCSS	
	All Participants			TNBC (n=93)		TNBC	
	5-yr OS Rate, %[1]	HR (95% CI)[2]	*P*	HR (95% CI)[3]	*P*	HR (95% CI)[4]	*P*
ALDH1A1							
Continuous	81.0	0.88 (0.75, 1.03)	0.11	1.12 (0.68, 1.84)	0.65	0.89 (0.81, 0.98)	0.02
<median	77.5	Reference		Reference		Reference	
≥median	84.6	0.84 (0.53, 1.34)	0.47	3.35 (0.70, 16.14)	0.13	0.65 (0.41, 1.04)	0.07
Tertile1	75.5	Reference		Reference		Reference	
Tertile 2	82.0	0.70 (0.42, 1.17)	0.17	1.68 (0.29, 9.71)	0.56	0.70 (0.41, 1.17)	0.17
Tertile 3	85.7	0.62 (0.34, 1.12)	0.11	1.87 (0.25, 14.15)	0.54	0.45 (0.25, 0.80)	0.007
GSTP1							
Continuous	81.0	0.93 (0.79, 1.10)	0.38	0.80 (0.43, 1.48)	0.47	1.07 (0.85, 1.33)	0.57
<median	82.4	Reference		Reference		Reference	
≥median	79.8	0.93 (0.56, 1.52)	0.76	1.12 (0.28, 4.44)	0.87	1.30 (0.83, 2.04)	0.26
Tertile1	77.9	Reference		Reference		Reference	
Tertile 2	84.3	0.87 (0.49, 1.54)	0.62	0.67 (0.10, 4.70)	0.69	1.16 (0.66, 2.04)	0.60
Tertile 3	80.4	0.87 (0.46, 1.66)	0.67	1.31 (0.19, 9.14)	0.79	1.20 (0.68, 2.09)	0.53
GSTA1							
Not Expressed	80.6	Reference		Reference		Reference	
Expressed	81.3	0.87 (0.54, 1.39)	0.56	0.46 (0.10, 2.13)	0.32	1.20 (0.57, 2.50)	0.63
GSTM1							
Not Expressed	85.5	Reference		Reference		Reference	
Expressed	80.3	1.29 (0.69, 2.41)	0.43	3.68E6 (0.00, .)	0.99	1.16 (0.73, 1.84)	0.52
GSTT1							
Not Expressed	78.1	Reference		Reference		Reference	
Expressed	81.7	0.77 (0.42, 1.44)	0.42	1.04 (0.15, 7.10)	0.97	1.13 (0.72, 1.78)	0.58

[1]Unadjusted, mean(se)
[2]Adjusted for age at diagnosis, TNM Stage, ER, PR, and HER2 status
[3]Adjusted for age at diagnosis and TNM stage
[4]Adjusted for age at diagnosis and tumor grade

Table C8: Overall Survival by Additive Cyclophosphamide Gene Scores in TCGA Breast Cancer Cases

		TCGA					SBCSS	
		All Subtypes			TNBC		TNBC	
		N	HR (95% CI)[2]	*P*	HR (95% CI)[2]	*P*	HR (95% CI)[2]	*P*
Score[4]								
	Group 1	243	Reference		Reference		Reference	
	Group 2	247	0.69 (0.37, 1.26)	0.23	0.11 (0.01, 2.22)	0.15	1.39 (0.81, 2.39)	0.24
	Group 3	397	0.88 (0.49, 1.58)	0.67	0.36 (0.03, 3.92)	0.40	1.26 (0.69, 2.28)	0.45
	P_{trend}		0.82		0.94		0.41	

[1]Adjusted for age at diagnosis, TNM stage, ER, PR, and HER2 status

[2]Adjusted for age at diagnosis and TNM stage

[3]Adjusted for age at diagnosis and tumor grade

[6]Score based on median cut point (higher score indicates faster metabolizer)

Table C9: Association Between Cyclophosphamide Gene Expression Levels and Scores Stratified by Early vs. Late Events in TCGA Breast Cancer Cases

	TCGA				SBCSS			
	Events <3 years (33 events)		Events ≥3 years (54 events)		Events <3 years (34 events)		Events ≥3 years (42 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*
ALDH1A1								
Continuous	0.81 (0.63, 1.04)	0.10	0.96 (0.77, 1.19)	0.71	1.10 (0.93, 1.32)	0.27	0.86 (0.75, 0.98)	0.03
<median	Reference		Reference		Reference		Reference	
≥median	0.55 (0.26, 1.15)	0.11	1.19 (0.64, 2.20)	0.58	1.22 (0.57, 2.61)	0.61	0.65 (0.34, 1.27)	0.21
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.53 (0.23, 1.24)	0.15	0.86 (0.44, 1.68)	0.66	1.16 (0.53, 2.58)	0.71	0.63 (0.29, 1.33)	0.22
Tertile 3	0.27 (0.10, 0.76)	0.01	1.14 (0.53, 2.43)	0.73	0.82 (0.30, 2.24)	0.70	0.50 (0.22, 1.11)	0.09
GSTP1								
Continuous	1.05 (0.80, 1.38)	0.71	0.80 (0.64, 1.01)	0.06	0.86 (0.59, 1.25)	0.43	1.01 (0.73, 1.41)	0.93
<median	Reference		Reference		Reference		Reference	
≥median	1.54 (0.65, 3.60)	0.32	0.58 (0.30, 1.14)	0.11	1.04 (0.50, 2.17)	0.91	1.07 (0.57, 2.03)	0.83
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.43 (0.14, 1.32)	0.14	1.19 (0.57, 2.51)	0.64	0.86 (0.32, 2.34)	0.77	0.96 (0.41, 2.23)	0.92
Tertile 3	1.59 (0.56, 4.51)	0.38	0.48 (0.20, 1.15)	0.10	0.86 (0.32, 2.31)	0.77	0.82 (0.36, 1.90)	0.65
GSTA1								
<median	Reference		Reference		Reference		Reference	
≥median	0.57 (0.27, 1.24)	0.16	1.16 (0.63, 2.16)	0.63	0.87 (0.30, 2.53)	0.80	0.96 (0.33, 2.77)	0.94
GSTM1								
<median	Reference		Reference		Reference		Reference	
≥median	0.85 (0.36, 1.97)	0.70	1.67 (0.68, 4.08)	0.26	1.67 (0.84, 3.30)	0.14	0.94 (0.49, 1.80)	0.86
GSTT1								
<median	Reference		Reference		Reference		Reference	
≥median	0.71 (0.26, 1.91)	0.49	0.72 (0.32, 1.62)	0.43	2.18 (1.08, 4.42)	0.03	0.79 (0.42, 1.51)	0.48
Score[2]								
Group 1	Reference		Reference		Reference		Reference	
Group 2	0.83 (0.32, 2.14)	0.70	0.63 (0.28, 1.42)	0.26	2.03 (0.81, 5.10)	0.13	1.14 (0.57, 2.25)	0.71
Group 3	0.50 (0.18, 1.36)	0.18	1.05 (0.49, 2.25)	0.90	2.68 (1.08, 6.65)	0.03	0.64 (0.26, 1.54)	0.32

[1]Adjusted for age at diagnosis, TNM stage, ER, PR, and HER2 status
[2]Adjusted for age at diagnosis and TNM stage
[3]Adjusted for age at diagnosis, TNM stage, ER, PR, and HER2 status
[2]Score based on median cut point (higher score indicates faster metabolizer)

5-Fluorouracil

The association between each SNP in genes involved in metabolism of 5-fluorouracil and OS among all breast cancer cases in the TCGA were evaluated (Table C10). The alleles were coded in the same way as for the SBCSS data for ease of comparison of the results. After adjustment for age at diagnosis, TNM stage, ER, PR, and HER2 status, the T allele in the SNP rs1801159 in the DPYD gene was significantly associated with better overall survival (HR=0.68, 95% CI: 0.48, 0.96, p=0.03) and the G allele in the SNP rs1801131 in the MTHFR gene was marginally associated with poorer overall survival (HR=1.36, 95% CI: 0.99, 1.86, p=0.06). No other associations between SNPs involved in cyclophosphamide metabolism and overall survival among TCGA breast cancer cases were observed. Neither of the observed associations was observed in the SBCSS data.

Each of the five gene scores created for SNPs in genes involved in cyclophosphamide metabolism were evaluated for their association with DFS and OS (Table C11). The deactivation gene score was associated with significantly better overall survival (HR=0.79, 95% CI: 0.65, 0.97, p=0.02. Both of the total scores were associated with significantly worse overall survival (p=0.02 for both). These results are opposite of what was hypothesized and observed in the SBCSS data.

When the results were stratified by the timing of survival events, those that occurred in the first three years of follow-up and those that occurred after three years, no clear differences were observed between any of the scores among those who had events in the first three years compared to those who had later events (Table C12).

Table C10: Associations Between Individual SNPs in 5-Fluorouracil Metabolizing Genes and Overall Survival in TCGA Breast Cancer Cases

		TCGA				SBCSS			
Gene	**rs ID**	**Allele Frequency**	**N**	**HR (95% CI)[1]**	***P***	**Allele Frequency**	**N**	**HR (95% CI)[2]**	***P***
Activation									
TYMP	rs11479	A=0.11	885	0.99 (0.56, 1.75)	0.97	A=0.21	3736	0.98 (0.83, 1.16)	0.84
UMPS	rs1801019	C=0.17	885	1.21 (0.82, 1.80)	0.34	C=0.18	3736	0.97 (0.82, 1.15)	0.71
UMPS	rs3772809	A=0.99	885	1.68 (0.23, 12.35)	0.61	A=0.94	3739	1.20 (0.89, 1.62)	0.22
Deactivation									
DPYD	rs17376848	G=0.05	885	0.54 (0.17, 1.69)	0.29	G=0.10	1124	1.20 (0.75, 1.93)	0.45
DPYD	rs1801159	T=0.81	885	0.68 (0.48, 0.96)	0.03	T=0.73	3736	0.98 (0.85, 1.13)	0.81
DPYD	rs1801265	A=0.75	885	0.83 (0.61, 1.14)	0.25	A=0.91	3736	1.06 (0.84, 1.34)	0.63
DPYD	rs72728438	T=0.81	885	0.97 (0.68, 1.37)	0.84	T=0.77	1124	1.28 (0.92, 1.77)	0.15
Other									
MTHFR	rs1801131	G=0.28	885	1.36 (0.99, 1.86)	0.06	G=0.18	3739	0.89 (0.75, 1.07)	0.22
MTHFR	rs1801133	G=0.69	885	1.11 (0.80, 1.55)	0.54	G=0.56	1124	0.93 (0.69, 1.24)	0.60
MTHFR	rs2274976	C=0.95	885	0.82 (0.39, 1.71)	0.59	C=0.91	3736	0.98 (0.78, 1.22)	0.85
TYMS	rs2847153	A=0.22	885	1.00 (0.70, 1.44)	1.00	A=0.36	1145	0.90 (0.69, 1.18)	0.46
TYMS	rs2853533	G=0.80	885	0.78 (0.51, 1.18)	0.24	G=0.49	1124	0.99 (0.75, 1.32)	0.95

[1]Adjusted for age at diagnosis, tumor grade, ER, PR, and HER2 status

[2]Adjusted for age at diagnosis, tumor grade, ER, PR, and HER2 status

Table C11: Association Between 5-Fluorouracil Gene Scores and Overall Survival in TCGA Breast Cancer Cases

		TCGA			SBCSS		
Role in Metabolism	**# of SNPs**	**N**	**HR (95% CI)[4]**	***P***	**N**	**HR (95% CI)[5]**	***P***
Activation[1]	3	884	1.15 (0.83, 1.58)	0.40	3736	1.00 (0.90, 1.11)	0.96
Deactivation[2]	4	884	0.79 (0.65, 0.97)	0.02	1124	1.01 (0.84, 1.21)	0.95
Other[3]	5	884	1.08 (0.89, 1.31)	0.42	1124	0.91 (0.76, 1.09)	0.29
Total Score[1,2]	7	884	1.22 (1.03, 1.44)	0.02	1124	0.91 (0.78, 1.05)	0.19
Total Score[1,2,3]	12	884	1.16 (1.02, 1.31)	0.02	1124	0.90 (0.80, 1.01)	0.08

[1]Includes SNPs from TYMP and UMPS genes
[2]Includes SNPs from DPYD gene
[3]Includes SNPs from TYMS and MTHFR genes
[4]Adjusted for age at diagnosis, TNM stage, and ER, PR, and HER2 status
[5]Adjusted for age at diagnosis, tumor grade, and ER, PR, and HER2 status

Table C12: Association Between 5-Fluorouracil Gene Scores and Overall Survival Stratified by Early vs Late Events in TCGA Breast Cancer Cases

		TCGA				SBCSS	
		Event <3 years		**Event ≥3 years**		**Event <3 years**	**Event ≥3 years**
Role in Metabolism	**# of SNPs**	**Events/ Total N**	**HR (95% CI)[4]**	**Events/ Total N**	**HR (95% CI)[4]**	**HR (95% CI)[5]**	**HR (95% CI)[5]**
Activation[1]	3	34/885	1.34 (0.80, 2.23)	59/277	0.98 (0.65, 1.49)	0.98 (0.84, 1.15)	1.02 (0.89, 1.18)
Deactivation[2]	4	34/885	0.66 (0.47, 0.91)	59/277	0.89 (0.69, 1.15)	0.90 (0.66, 1.21)	1.07 (0.85, 1.35)
Response[3]	5	34/885	0.81 (0.59, 1.10)	59/277	1.29 (1.00, 1.66)	0.94 (0.70, 1.26)	0.89 (0.71, 1.12)
Total Score[1,2]	7	34/885	1.45 (1.10, 1.90)	59/277	1.07 (0.88, 1.32)	1.05 (0.83, 1.33)	0.83 (0.69, 1.00)
Total Score[1,2,3]	12	34/885	1.13 (0.91, 1.40)	59/277	1.14 (0.98, 1.32)	1.00 (0.83, 1.22)	0.84 (0.73, 0.98)

[1]Includes SNPs from TYMP and UMPS genes
[2]Includes SNPs from DPYD gene
[3]Includes SNPs from TYMS and MTHFR genes
[4]Adjusted for age at diagnosis, TNM stage, ER, PR, and HER2 status
[5]Adjusted for age at diagnosis, tumor grade, and ER, PR, and HER2 status

The gene expression levels of each of the genes involved in 5-fluorouracil metabolism among TCGA breast cancer cases are shown in Table C13. All genes of interest were expressed in all of the tumor tissues from TCGA breast cancer cases.

Table C13: Expression Levels of 5-Fluorouracil Metabolizing Genes in TCGA Breast Cancer Cases

Gene	% Expressed	Mean	SD	Median	Minimum	Maximum
DPYD	100.0%	8.52	1.07	8.63	4.42	11.59
TYMS	100.0%	9.12	1.08	9.09	5.85	12.76
MTHFR	100.0%	9.29	0.83	9.30	6.49	11.63
TYMP	100.0%	10.45	1.18	10.47	5.66	13.62
UMPS	100.0%	9.55	0.42	9.52	8.18	10.99

The gene expression levels for genes involved in metabolism of 5-fluorouracil and overall survival among all TCGA breast cancer cases and TCGA TNBC cases only were evaluated (Table C14). TYMP expression was associated with significantly improved survival among all breast cancer cases (continuous analysis HR=0.82, 95% CI: 0.70, 0.96, p=0.02). This is similar in magnitude and direction to the SBCSS results when the analysis was restricted to those who took 5-fluorouracil (HR=0.79, 95% CI: 0.51, 1.23), though the SBCSS finding was not significant. No association between TYMP and OS was observed among TCGA TNBC cases only. When analyzed by tertiles, UMPS expression was associated with significantly worse overall survival among TCGA TNBC cases (highest tertile compared to lowest HR=14.63, 95%

CI: 1.19, 179.43, p=0.04), but not among all breast cancer cases. No other significant associations were observed between any of the other gene expression levels and survival in the TCGA, either among all breast cancer cases or TNBC only.

No association was observed between the additive 5-fluorouracil gene score and OS among all breast cancer cases (Table C15). There were too few events among TNBC cases to run this analysis.

When stratified by events in the first three years and events occurring after three years, DPYD expression was associated with better survival in the first 3 years following diagnosis (HR based on median cut point=0.44, 95% CI: 0.20, 0.95, p=0.04) and worse survival after three years (HR based on median cut point=2.38, 95% CI: 1.31, 4.32, p=0.004) (Table C16). TYMP expression was associated with better survival only among those who had events after three years (continuous HR=0.73, 95% CI: 0.59, 0.89, p=0.002). UMPS was significantly associated with worse survival among those who had an event in the first three years (continuous HR=3.52, 95% CI: 1.48, 8.33, p=0.004) but not those who had an event after three years. This finding is similar in magnitude and direction as the SBCSS when analyzed by median cut point (TCGA: HR=2.39, 95% CI: 1.10, 5.18, p=0.03; SBCSS: HR=2.14, 95% CI: 1.03, 4.43, p=0.04).

The 5-fluorouracil gene score was associated with significantly better overall survival among those who had events in the first three years after diagnosis (highest score compared to lowest score HR=0.10, 95% CI: 0.01, 0.97, p=0.05) and significantly worse survival among those who had events after three years (highest score compared to lowest score HR=3.15, 95% CI: 1.08, 9.20, p=0.04). In the SBCSS, a similar difference in survival outcomes was observed, although the estimates were not significant and no significant difference was detectable.

Table C14: Overall Survival by Expression of Genes Involved in 5-Fluorouracil Metabolism in TCGA Breast Cancer Cases

	TCGA					SBCSS	
	All Participants			TNBC (n=93)		TNBC	
	5-yr OS Rate, %[1]	HR (95% CI)[2]	*P*	HR (95% CI)[3]	*P*	HR (95% CI)[4]	*P*
DPYD							
Continuous	81.0	1.00 (0.82, 1.23)	0.98	1.40 (0.82, 2.38)	0.21	0.92 (0.82, 1.04)	0.18
<median	80.9	Reference		Reference		Reference	
≥median	80.5	1.32 (0.83, 2.09)	0.24	6.26 (1.13, 34.75)	0.04	0.70 (0.44, 1.10)	0.12
Tertile 1	79.0	Reference		Reference		Reference	
Tertile 2	84.0	1.08 (0.65, 1.81)	0.76	9.90 (0.82,119.99)	0.07	0.62 (0.36, 1.07)	0.09
Tertile 3	79.6	1.10 (0.61, 2.01)	0.75	7.48 (0.67, 83.52)	0.10	0.70 (0.41, 1.21)	0.20
TYMS							
Continuous	81.0	0.97 (0.78, 1.20)	0.78	0.63 (0.31, 1.28)	0.20	1.09 (0.95, 1.24)	0.24
<median	79.1	Reference		Reference		Reference	
≥median	82.1	0.89 (0.56, 1.41)	0.61	0.10 (0.01, 0.77)	0.03	1.55 (0.97, 2.45)	0.06
Tertile 1	75.7	Reference		Reference		Reference	
Tertile 2	86.3	0.59 (0.34, 1.03)	0.06	0.81 (0.17, 3.91)	0.80	1.17 (0.65, 2.11)	0.60
Tertile 3	79.2	0.94 (0.52, 1.69)	0.83	0.06 (0.00, 0.93)	0.04	1.73 (0.99, 3.02)	0.06
MTHFR							
Continuous	81.0	0.82 (0.63, 1.08)	0.16	1.43 (0.73, 2.77)	0.29	0.91 (0.81, 1.03)	0.13
<median	79.0	Reference		Reference		Reference	
≥median	83.1	0.98 (0.62, 1.54)	0.93	1.46 (0.32, 6.65)	0.63	0.62 (0.39, 0.98)	0.04
Tertile 1	75.9	Reference		Reference		Reference	
Tertile 2	85.8	1.48 (0.85, 2.56)	0.16	6.62 (0.91, 48.28)	0.06	0.76 (0.46, 1.27)	0.30
Tertile 3	82.4	0.87 (0.46, 1.62)	0.65	1.10 (0.17, 7.27)	0.92	0.42 (0.23, 0.76)	0.004
TYMP							
Continuous	81.0	0.82 (0.70, 0.96)	0.02	1.02 (0.57, 1.83)	0.94	1.02 (0.81, 1.30)	0.85
<median	78.5	Reference		Reference		Reference	
≥median	83.6	0.51 (0.32, 0.82)	0.005	1.57 (0.36, 6.91)	0.55	1.18 (0.75, 1.85)	0.48
Tertile 1	78.7	Reference		Reference		Reference	
Tertile 2	79.4	1.04 (0.62, 1.72)	0.89	2.24 (0.38, 13.34)	0.37	1.14 (0.67, 1.95)	0.63
Tertile 3	85.6	0.41 (0.23, 0.74)	0.003	3.58 (0.48, 26.42)	0.21	0.96 (0.54, 1.71)	0.90

UMPS							
Continuous	81.0	1.52 (0.90, 2.57)	0.12	2.55 (0.54, 12.07)	0.24	0.99 (0.89, 1.11)	0.92
<median	82.7	Reference		Reference		Reference	
≥median	79.3	1.40 (0.90, 2.18)	0.13	1.99 (0.34, 11.54)	0.44	1.39 (0.88, 2.20)	0.16
Tertile 1	81.6	Reference		Reference		Reference	
Tertile 2	88.2	0.81 (0.44, 1.49)	0.49	1.17 (0.13, 10.64)	0.89	0.94 (0.53, 1.66)	0.83
Tertile 3	73.7	1.23 (0.72, 2.10)	0.44	14.63 (1.19,179.43)	0.04	1.31 (0.76, 2.28)	0.33

[1]Unadjusted, mean(se)
[2]Adjusted for age at diagnosis, TNM Stage, ER, PR, and HER2 status
[3]Adjusted for age at diagnosis and TNM stage
[4]Adjusted for age at diagnosis and tumor grade

Table C15: Overall Survival by Additive 5-Fluorouracil Gene Scores in TCGA Breast Cancer Cases

		TCGA			SBCSS	
		All Subtypes			TNBC	
		N	**HR (95% CI)[1]**	***P***	**HR (95% CI)[1]**	***P***
Score[4]						
	Group 1	132	Reference		Reference	
	Group 2	308	0.97 (0.47, 2.01)	0.94	0.97 (0.48, 1.96)	0.93
	Group 3	312	1.17 (0.56, 2.46)	0.67	0.85 (0.41, 1.74)	0.65
	Group 4	135	1.29 (0.57, 2.92)	0.54	0.67 (0.29, 1.57)	0.36
	P_{trend}		0.38		0.29	

[1]Adjusted for age at diagnosis, TNM stage, ER, PR, and HER2 status

[2]Adjusted for age at diagnosis and TNM stage

[3]Adjusted for age at diagnosis and tumor grade

[4]Score based on median cut point (higher score indicates faster metabolizer, slower activator, and poorer response)

Table C16: Association Between 5-Fluorouracil Gene Expression Levels and Scores Stratified by Early vs. Late Events in TCGA Breast Cancer Cases

	TCGA				SBCSS			
	Events <3 years (33 events)		Events ≥3 years (54 events)		Events <3 years (34 events)		Events ≥3 years (42 events)	
	HR (95% CI)[1]	*P*	HR (95% CI)[1]	*P*	HR (95% CI)[2]	*P*	HR (95% CI)[2]	*P*
DPYD								
Continuous	0.77 (0.58, 1.01)	0.06	1.28 (0.98, 1.68)	0.07	1.10 (0.89, 1.36)	0.39	0.92 (0.77, 1.10)	0.37
<median	Reference		Reference		Reference		Reference	
≥median	0.44 (0.20, 0.95)	0.04	2.38 (1.31, 4.32)	0.004	0.66 (0.32, 1.38)	0.27	0.97 (0.52, 1.81)	0.92
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.63 (0.28, 1.43)	0.27	1.78 (0.91, 3.49)	0.09	0.97 (0.20, 4.78)	0.97	0.88 (0.41, 1.90)	0.74
Tertile 3	0.28 (0.10, 0.79)	0.02	2.47 (1.15, 5.31)	0.02	1.40 (0.29, 6.69)	0.67	1.12 (0.51, 2.44)	0.78
TYMS								
Continuous	0.89 (0.63, 1.26)	0.53	0.96 (0.73, 1.27)	0.77	0.98 (0.81, 1.18)	0.80	0.99 (0.81, 1.20)	0.90
<median	Reference		Reference		Reference		Reference	
≥median	0.79 (0.38, 1.64)	0.52	0.97 (0.53, 1.78)	0.92	1.03 (0.44, 2.41)	0.95	1.06 (0.50, 2.27)	0.88
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.88 (0.35, 2.24)	0.79	0.42 (0.20, 0.87)	0.02	1.78 (0.64, 4.95)	0.27	0.48 (0.19, 1.19)	0.11
Tertile 3	0.96 (0.38, 2.43)	0.93	0.86 (0.41, 1.81)	0.68	1.20 (0.37, 3.86)	0.76	0.98 (0.37, 2.60)	0.97
MTHFR								
Continuous	0.93 (0.60, 1.42)	0.72	0.76 (0.53, 1.08)	0.12	1.10 (0.88, 1.37)	0.42	0.95 (0.78, 1.14)	0.56
<median	Reference		Reference		Reference		Reference	
≥median	1.39 (0.67, 2.88)	0.38	0.75 (0.42, 1.34)	0.33	0.88 (0.40, 1.92)	0.74	0.83 (0.42, 1.64)	0.60
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	1.51 (0.67, 3.40)	0.32	1.46 (0.68, 3.12)	0.33	1.02 (0.21, 5.01)	0.98	0.90 (0.43, 1.87)	0.77
Tertile 3	1.08 (0.40, 2.89)	0.88	0.70 (0.31, 1.62)	0.41	1.46 (0.31, 6.98)	0.63	0.57 (0.24, 1.35)	0.20
TYMP								
Continuous	0.95 (0.72, 1.25)	0.70	0.73 (0.59, 0.89)	0.002	0.92 (0.67, 1.26)	0.60	0.98 (0.71, 1.35)	0.90
<median	Reference		Reference		Reference		Reference	
≥median	0.77 (0.37, 1.61)	0.49	0.33 (0.17, 0.63)	0.0008	1.10 (0.55, 2.21)	0.78	0.98 (0.53, 1.80)	0.94
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	2.10 (0.88, 5.01)	0.10	0.62 (0.31, 1.21)	0.16	1.16 (0.51, 2.62)	0.72	1.13 (0.54, 2.35)	0.75
Tertile 3	0.82 (0.30, 2.25)	0.70	0.27 (0.12, 0.60)	0.001	1.00 (0.97, 1.04)	0.50	0.88 (0.40, 1.93)	0.75

UMPS								
Continuous	3.52 (1.48, 8.33)	0.004	1.02 (0.49, 2.12)	0.95	1.11 (0.91, 1.36)	0.31	0.94 (0.81, 1.09)	0.42
<median	Reference		Reference		Reference		Reference	
≥median	2.39 (1.10, 5.18)	0.03	1.11 (0.63, 1.98)	0.71	2.14 (1.03, 4.43)	0.04	1.10 (0.60, 2.02)	0.77
Tertile 1	Reference		Reference		Reference		Reference	
Tertile 2	0.82 (0.24, 2.84)	0.76	0.85 (0.41, 1.76)	0.66	1.18 (0.46, 3.07)	0.73	6.27 (0.83, 47.62)	0.08
Tertile 3	3.16 (1.17, 8.49)	0.02	0.74 (0.35, 1.56)	0.44	2.07 (0.86, 4.99)	0.10	4.47 (0.57, 34.79)	0.15
Score[3]								
Group 1	Reference		Reference		Reference		Reference	
Group 2	1.38 (0.45, 4.21)	0.58	0.57 (0.20, 1.66)	0.30	2.45 (0.71, 8.50)	0.16	0.38 (0.14, 1.04)	0.06
Group 3	0.68 (0.19, 2.40)	0.55	1.79 (0.66, 4.88)	0.25	0.75 (0.19, 3.03)	0.69	0.94 (0.40, 2.20)	0.89
Group 4	0.10 (0.01, 0.97)	0.05	3.15 (1.08, 9.20)	0.04	0.91 (0.21, 4.03)	0.91	0.65 (0.22, 1.89)	0.43

[1]Adjusted for age at diagnosis and TNM stage
[2]Adjusted for age at diagnosis, tumor grade, basal-like subtype, and DUSP4 expression
[3]Score based on median cut point (higher score indicates faster metabolizer, slower activator, and poorer response)