Genome-wide Analysis of Therapeutic Response Uncovers Molecular Pathways Governing Tamoxifen Resistance in ER+ Breast Cancer

By Sarra M. Rahem

A Dissertation Submitted

In partial fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biomedical Informatics

Department of Health Informatics

Rutgers, The State University of New Jersey

School of Health Professions

August 2020

Dissertation Approval Form

Genome-wide Analysis of Therapeutic Response Uncovers Molecular Pathways Governing Tamoxifen Resistance in ER+ Breast Cancer

BY

Sarra M. Rahem

Dissertation Committee:

Antonina Mitrofanova, PhD

Shankar Srinivasan, PhD

Frederick Coffman, PhD

Approved by the Dissertation Committee:

_____ **Date:** _____

_____ **Date:** _____

_____ **Date:** _____

_____ **Date:** _____

TABLE OF CONTENTS

ABSTRACT

Despite recent advances in diagnosis, classification, and therapeutic management, breast cancer (BC) remains one of the leading causes of cancer-related death in women worldwide. Nearly 70% of all diagnosed cases of breast tumors are Estrogen Receptor positive (ER+) and thus anti-estrogen therapy, such as tamoxifen, has become the standard-of-care for patients with ER+ breast cancers. Yet, nearly 30% of patients treated with tamoxifen develop resistance, ultimately leading to metastasis and lethality. Prioritization of breast cancer patients based on the risk of resistance to tamoxifen plays a significant role in personalized therapeutic planning and improving disease course and outcomes. In this work, we demonstrate that a genome-wide pathway-centric computational framework elucidates molecular pathways as markers of tamoxifen resistance in ER+ breast cancer patients. Through the association of pathway activity and response to tamoxifen, we identified five biological pathways and demonstrated their ability to predict the risk of tamoxifen resistance in two independent patient cohorts (Test cohort1: log-rank p-value = 0.02, adjusted HR = 3.11; Test cohort2: log-rank p-value = 0.01, adjusted HR = 4.24). Importantly, as a negative control, we have demonstrated that the identified 5 candidate pathways did not classify patients simply based on the disease aggressiveness and that pathways of aggressiveness do not overlap with the 5 candidate pathways. Finally, we have compared our pathway signature to other known signatures of tamoxifen response and have shown superiority of our pathway-based approach (adjusted hazard ratio = 3.11, hazard p-value=0.0278). Thus, we propose that the identified pathways as well as their representative read-out-genes can be utilized to prioritize patients who would benefit from

tamoxifen treatment and patients at risk of tamoxifen resistance that should be offered alternative regimens.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my dissertation committee chair Dr. Antonina Mitrofanova for her continuous support and guidance of my PhD research project. Thank you for encouraging my research and for allowing me to grow as a research scientist. This dissertation would not have been completed without your assistance. I would like to thank my committee members Dr. Frederick Coffman, Dr. Shankar Srinivasan, and Dr. Riddhi Vyas for serving as my committee members and providing me with invaluable input and feedback throughout the course of the dissertation process.

I would also like to thank all my lab members for their useful discussions and comments which helped me to improve the quality of my work. Moreover, I really would like to thank all the faculty and staff members of the Department of Health Informatics at Rutgers Biomedical and Health Sciences for their support and kindness.

I am very thankful to the people, who have made me who I am today, my mother Fatma Abuain and my father Mahfoud Rahem for their endless love, support, encouragement, and confidence in me. Finally, I would like to give special thanks to my son Hamza and my sisters Ghada, Rema, Esra and Mawada, whose support and love kept me going to make this work possible.

LIST OF TABELS

LIST OF FIGURES

Chapter I

INTRODUCTION

1.1 Background, Literature Review and Statement of the Problem

Despite recent advances in diagnosis, classification, and therapeutic management, breast cancer (BC) remains one of the leading causes of cancer-related death in women worldwide.[1-3] The main molecular subtypes of BC are based on the gene expression profiling, including luminal A, luminal B, HER2-enriched, and triple-negative/basal-like (Table 1). [1] The luminal type of breast cancer, which is mainly categorized by estrogen receptor-positive (ER+), is morphologically well-differentiated and displays a comparatively good prognosis compared with ER^- breast cancers which tend to be poorly differentiated and display a poor prognosis. [1] However, the studies have shown that 65–70% of breast cancers are luminal A and B tumors, whereas about 10% of breast cancers are HER2-enriched tumors and 10–19 % of breast cancers are basal-like tumors.[1] Those molecular classifications of BC subtypes significantly enhanced our understanding of the complicated properties of various breast tumors, their clinical overall outcomes, and their treatment responses. [1]

Table 1.

Different Characteristics of breast Cancer subtypes.

Breast Cancer Molecular Subtypes	
Subtypes	Characteristics
Luminal A	ER+ and/or PR+, HER2- and Ki-67 ≤14%
Luminal B	ER+ and/or PR+, HER2+/- and Ki-67 >14%/any Ki-67
HER2 –Enriched/Non-Luminal	ER-, PR-, HER2+ and any Ki-67
Triple Negative/Basal-Like	ER-, PR-, HER2- and any Ki-67

Nearly 70% of all diagnosed cases of breast tumors are ER+,[4,5] making treatments with anti-estrogen effects in the breast cells, such as tamoxifen, the standard-of-care for patients with ER+ breast cancers.[4,6-9] Despite the significant success of tamoxifen administration, nearly 30% of treated patients develop therapeutic resistance, ultimately leading to metastasis and lethality.[1,10] Therefore, prioritization of patients based on the risk of resistance to tamoxifen before treatment administration could play a significant role in personalize therapeutic planning for patients with ER+ breast cancer and builds a foundation to improve disease course and outcomes. In addition, identifying which patients are not likely to develop tamoxifen resistance is equally as important as identifying which patients are at high risk for developing tamoxifen resistance.

Tamoxifen is a selective estrogen receptor modulator (SERM) and has agonist or antagonist activity depending on the tissue type.[11] In the breast cells, tamoxifen directly binds to the ER, blocking estrogen from attaching to the receptor and thus inhibiting the activity of estrogen-regulated genes and causing the repression of estrogenic effects.[4,5,12,13] However, the emergence of alternative mechanisms of estrogenic stimulation has been shown to cause emergence of resistance to tamoxifen. For example, some studies have demonstrated that ER+ breast cancers that overexpress HER2 and EGFR can activate the components of downstream signaling pathways which then stimulate both ER and estrogen receptor co-activator AIB1, and thus induce the estrogen agonistic activity of tamoxifen in breast cancer cells.[14,15] Another study noticed that the increased expression of HER2 signaling can also downregulate progesterone receptor (PR) levels in the ER+ breast tumors, where losing the PR expression serves as a biomarker of hyperactive growth factor signaling, leading to another possible mechanism of tamoxifen resistance.[16] Despite the emerging role of HER2 in tamoxifen resistance, it only accounts for 10% of ER+ breast cancers,[12,17] suggesting more complex resistance mechanisms in these cases, presenting a central clinical problem for patients with ER+ breast cancer.[4,5,10,12]

In recent years, several groups have developed gene expression signatures of tamoxifen response for ER+ patients, including 10 gene-signature by Men et al.,[18] 21 gene-signature by Paik et al.[19] (known as Oncotype DX) and 2 gene-signature by Ma et al.[20] While these signatures provide substantial advances to our understanding of individual genes involved in resistance, they do not yet capture the complex interplay between biological mechanisms that governs tamoxifen resistance. Here we propose a pathway-centric computational framework to elucidate tamoxifen resistance and demonstrate that it

outperforms known gene-based approaches. Advantages of our pathway-based approach lies in *(i)* its ability to identify a tightly connected cooperative group of genes unified by the same function;[21-23] *(ii)* studying molecular pathways, rather than individual genes, produces more reliable read-out outputs as they are less susceptible to experimental noise;[24] *(iii)* pathway-level view enhances our understanding of the biological mechanisms related to disease and treatment response;[25-28] and finally *(iv)* looking at alterations in biological pathways enhances the likelihood of identifying potential therapeutic targets to preclude or overcome resistance.

1.2 Research Hypotheses

We suggest that the defined candidate pathways as well as their representative read-out-genes can potentially be used to identify patients who would benefit from the tamoxifen treatment as their first-line therapy and those at risk of developing therapy failure, even prior to treatment administration, which enhances the personalized and precision treatment strategy. In fact, identifying which patients are not likely to develop tamoxifen resistance is equally as important as identifying which patients are at high risk for developing tamoxifen resistance. While the identified molecular pathways act as promising predictive markers for treatment response, they can also be potential candidates for therapeutic target to prevent resistance. Although this work is focused on identifying patients with high potential to antiestrogen resistance, our approach can be broadly applicable to other therapeutic interventions and diseases.

1.3 An Overview of the Study

In this work, we have established a systematic pathway-centric computational framework to elucidate molecular pathways as markers of tamoxifen resistance in ER+ breast cancer patients. Through the analysis of pathway activity in each ER+ patient and their association with response to tamoxifen (n = 53), we identified five biological pathways as pathways essential for tamoxifen resistance: Retrograde Neurotrophin Signalling, Loss of NLP from Mitotic Centrosomes, RNA Polymerase III Transcription Initiation from Type 2 Promoter, EIF2 pathway, and Valine, Leucine and Isoleucine Biosynthesis. We have demonstrated the ability of the identified five (5) candidate pathways to predict the risk of tamoxifen resistance in two independent patient cohorts[29] (Test cohort 1, n = 66 : log-rank p-value = 0.02, accuracy in leave one out cross-validation (LOOCV) = 85.8%; Test cohort 2, n = 77: log-rank p-value = 0.01, accuracy in LOOCV = 82.5%) and their independence from known covariates, such as age, tumor grade, tumor size, lymph node status, and PR status, as the absence of PR in ER+ tumor can be an indicator of HER2 activation and an aggressive phenotype[16] (Test cohort 1, adjusted hazard ratio = 3.11; Test cohort 2, adjusted hazard ratio = 4.24). In addition, we have shown significant non-random predictive ability of our pathways, when compared to pathways chosen at random (random model p-value=0.031). Furthermore, we performed stratified Kaplan-Meier survival analysis, where we evaluated predictive ability of our candidates on patient groups divided by PR status, age groups, and luminal subtypes and demonstrated that the five candidate pathways can predict risk of resistance to tamoxifen in each group. Importantly, as a negative control, we have demonstrated that the identified five candidate pathways did not classify patients simply based on the disease aggressiveness (log-rank p-value = 0.7, hazard ratio = 1.246) and that in fact pathways associated with disease

aggressiveness do not overlap with the five candidate pathways. We have compared our method to other computational techniques to tackle treatment response, including Epsi et al.[27] (which utilized extreme-responder analysis, using tails of the treatment response distribution to define a treatment response signature), Zhong et al.[30] (which used Support Vector Machine approach as a base), and Yu et al.[31] (which uses random forest approach as a base) and demonstrated that our method outperforms these techniques in predicting risk of resistance to tamoxifen. Further, we have compared our pathway signature to other known signatures of tamoxifen response[18-20] and have shown the superiority of our pathway-based approach (adjusted hazard ratio = 3.11, hazard p-value = 0.0278). Finally, we identified 5 read-out genes that can function as biomarker to identify patients at risk of developing resistance to tamoxifen in testing cohort. Thus, we propose that the identified five candidate pathways and their corresponding read-out genes can potentially be used to prioritize patients who would benefit from tamoxifen treatment as their first-line therapy, and to identify patients at risk of tamoxifen resistance who should be offered an alternative regimen plan.

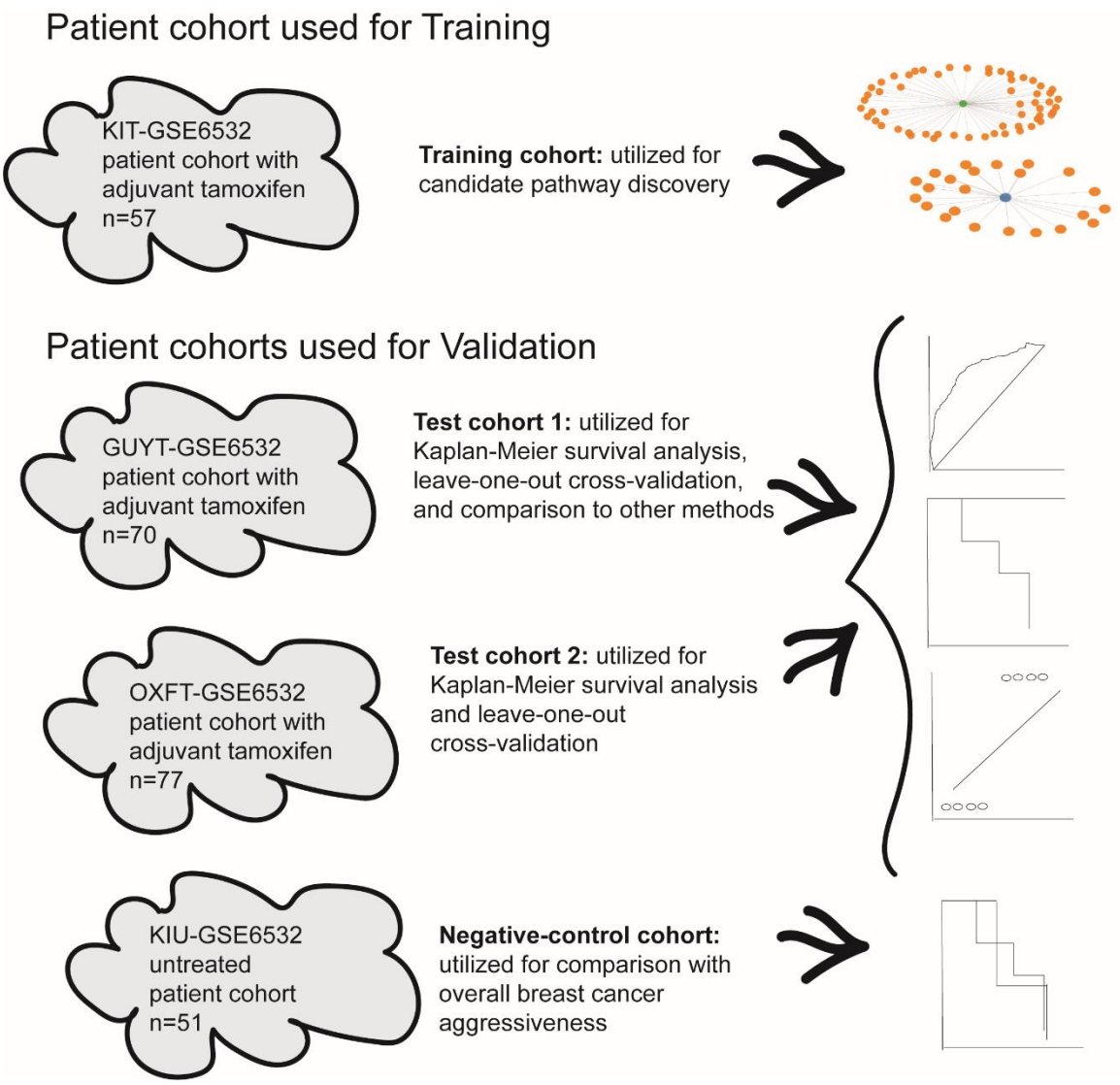
Chapter II

METHODS

2.1 Patient cohorts utilized in this study

All gene expression datasets of patients with ER+ breast cancer were obtained from publicly available GEO data repository[32] from multi-institutional multi-PI comprehensive Loi et al.[29] study GSE6532 (Figure 1, Table 2): *(i)* KIT-GSE6532 utilized as a Training cohort; *(ii)* GUYT-GSE6532, utilized as Test cohort 1; *(iii)* OXFT-GSE6532, utilized as a Test cohort 2; and *(iv)* KIU-GSE6532, utilized as a negative control cohort. Training cohort contains patient profiles of primary ER+ breast tumors (n = 57), archived at the Uppsala University Hospital (Uppsala, Sweden), profiled on Affymetrix Human Genome U133A array and Affymetrix Human Genome U133B array. Test cohort 1 contains patient profiles of primary tumors from patients with ER+ breast cancer (n = 70), archived at the Guy's Hospital (London, United Kingdom), profiled on Affymetrix Human Genome U133 Plus 2.0 Array. Test cohort 2 contains patient profiles of primary ER+ breast tumors (n = 77), archived at the John Radcliffe Hospital (Oxford, United Kingdom), profiled on Affymetrix Human Genome U133A, B array. Negative control cohort consists of not-treated patients with ER+ primary breast tumors (n = 51), profiled on Affymetrix Human Genome U133A, B array. All primary tumors samples in Training and Test cohorts collected through surgery, diagnosed between 1980 and 1995 and received tamoxifen-only treatment for 5 years post-diagnosis as their adjuvant treatment.

Figure 1.



Schematic representation of the utilization of independent patient cohorts for Training, Testing, and negative-control purposes utilized in this study.

Table 2.

Clinical characteristics of datasets used for Training and Testing analysis

Characteristics	**Training cohort KIT-GSE6532**	**Test cohort 1 GUYT-GSE6532**	**Test cohort 2 OXFT-GSE6532**	**Negative-control cohort KIU-GSE6532**
Platform	Affymetrix Human Genome U133A, B array	Affymetrix Human Genome U133 Plus 2.0 Array	Affymetrix Human Genome U133A, B array	Affymetrix Human Genome U133A, B array
Total number of patients	57	70	77	51
PAM50 Classification				
Luminal A	45	27	61	43
Luminal B	8	39	16	4
*Other subtypes	4	4	-	4
Number of patients utilized in our study (other subtypes excluded)	53	66	77	47
Number of events	22/53 (41.5%)	19/66 (28.78%)	20/77 (25.97%)	17/47 (36.17%)
Age				
=<50	1	3	13	19
>50	52	63	64	28
Histological grade				
1	10	17	17	22
2	37	36	46	20
3	6	13	14	5

Tumor size				
≤2 cm	19	35	35	26
>2 cm	34	31	42	21
Lymph node status				
Negative	14	22	48	47
Positive	39	44	29	-
PR status				
Negative	4	16	-	1
Positive	49	50	-	46

Note: * other subtypes = HER2-enriched, basal-like, or normal-like.

2.2 Data normalization

For each gene expression microarray dataset, matrix of RMA (Robust Microarray Analysis) normalized signal intensity values were used.[48] Using the current annotation file from GEO and the latest Affymetrix annotation files from Thermo Fisher database,[33] each probe set ID was annotated to gene ID, thereafter, probe IDs that annotated to different gene IDs or did not annotate to any gene ID were excluded. When multiple probe set IDs were mapped to the same gene, probes with the highest coefficient variation (CV) over all samples were selected. The CV for each probe was computed by dividing the standard deviation of expression values among all sample population by the mean expression value.[34,35]

2.3 Determining the molecular subtypes of breast cancer patients

Gene expression classifier (PAM50) of the breast cancer subtypes was applied to assign BC patients to one of the intrinsic molecular subtypes: luminal A, luminal B, HER2-enriched, triple-negative/basal-like, and normal-like.[36,37] The subtype classification of each patient was determined based on the closeness between the average expression profile of 50 genes in each subtype centroid and the corresponding gene expression pattern of patient tumor, where the distances measured utilizing Spearman's rank correlation.[36] From *genefu* package in R, *intrinsic.cluster.predict* function with pam50[38] was utilized to eliminate samples with HER2-enriched, triple-negative/basal-like, and normal-like subtypes (i.e., non ER+).

2.4 Single-Sample Gene Set Enrichment Analysis (ssGSEA)

For single-sample analysis, gene expression values for each gene were transformed into standardized scores (i.e., z-scores) in order to bring the expression level into a common scale across all samples.[39,40] Z-score for each gene was computed by subtracting the average intensity of the gene from the intensity of this gene in each sample and dividing it by the standard deviation (SD) across all samples.[39] In this way, each gene's mean is standardized to 0 and standard deviation to 1. Ranked list of z-scored for each gene for a given sample then defines a single-sample signature, utilized for further pathway enrichment analysis.

For pathway enrichment analysis, we utilized Reactome,[41] BioCarta[42] and KEGG[43] databases, which contains 833 biological pathways, and implemented single-sample GSEA (i.e., ssGSEA[44,45]), where each single-sample signature was used as a reference,

and each pathway (i.e., genes from each pathway) was used as a query set. The GSEA normalized enrichment scores (NESs), and p-values were assessed utilizing 1,000 gene permutations. NES for each of the 833 pathways (i.e., also referred to as pathway activity levels) indicated how much each pathway is overrepresented in each single-sample signature. In particular, the positive NES would indicate a pathways enrichment in the top of the rank-ordered list (i.e., overexpressed part) of the signature and the negative NES would indicate pathway enrichment in the bottom of the rank ordered list (i.e., underexpressed part) of the signature.

2.5 Associating the activity levels of molecular pathways with therapeutic response

The activity levels of each pathway (i.e., NES) were then associated with tamoxifen response using Cox proportional hazards model,[46] adjusted for common covariates, such as age, tumor grade, tumor size, lymph node status, and PR status. For this, we utilized R *coxph* function from *survival* package.[47] To establish a robust threshold which should be utilized to select most significantly associated pathways, we evaluated predictive ability of the pathways as a group (starting from the most significant pathway and then adding the next most significant pathway, one at a time). Thus, the groups of pathways that were evaluated were *(i)* Pathway 1; *(ii)* Pathways 1 and 2; *(iii)* Pathways 1, 2, and 3; etc. until all pathways were utilized. We then evaluated predicted ability of each group and recorded them (see Results). The cutoff point was determined as the one, where the addition of a pathway would not benefit an overall predictive ability of the group.

Furthermore, given that many of the 833 pathways exhibit parent-child relationships are heavily overlapping, we examined all final pathways that had the above relationships and if such dispute occurred, we prioritized pathways with higher association with tamoxifen response.

2.6 Clinical validation in independent patient cohorts

For validation studies, the activity levels of the five candidate pathways were used to stratify patients based on the risk of relapse due to treatment resistance in independent Test cohorts. Patient cohorts were subjected to t-SNE clustering, a widely-utilized dimensionality reduction technique,[48] using all pairs of high-dimensional (i.e., 5-dimensions in this study) points.[49,50] In fact, t-SNE reduces high-dimensional dataset (i.e., 5-dimensional) in a low-dimensional (i.e., 2-dimensional) space and successfully distinguishes groups of patients that have similar pathway activity levels. Subsequently, k-means clustering[51] was utilized on t-SNE-derived the low-dimensional (i.e., 2-dimensional) space to obtain two groups of patients with distinct pathway activity patterns,[49,50] using *kmeans* function in R.[52]

The ability of the activity levels of the 5 molecular pathways to efficiently distinguish patient clusters was determined through receiver operating characteristics (ROC) analysis[53] on multiple (i.e., multivariable) logistic regression model, where normalized enrichment scores of 5 pathways were used as input parameters (i.e., independent/predictor variables) and patient clusters were utilized as a dependent/response variable. ROC curves were assessed using the area under the curve (AUC),[54] where AUC

score of 0.5 indicates a random predictor. The logistic regression analysis was conducted using *glm*[55] function, and ROC analysis was performed using *pROC*[56] and *ggplot2* packages in R.

Differences in therapeutic response between the patient groups were evaluated through Kaplan-Meier treatment-related survival analysis[57] and Cox proportional hazards model using *survival* and *survminer* packages [46] in R. Log-rank p-value was utilized to assess the statistical significance of the Kaplan-Meier survival analysis and Wald p-value and hazard ratio were utilized for multivariable Cox proportional hazards model through *survdiff* and *coxph* functions from *survival* package.

To estimate the predictive accuracy of our model and obtain a more accurate indication of how well our finding behaves toward a new incoming patient, we conducted Leave-one-out cross-validation (LOOCV).[58] In this method, one patient is "excluded/eliminated" and the rest of the patients are utilized for training purposes to the regression model. After that, a removed patient is assumed to be a new incoming patient and is assigned a risk of developing tamoxifen resistance. This process is repeated for each patient within a given dataset. LOOCV was implemented for multiple logistic regression model, where patient clusters membership was used as a response variable and normalized enrichment scores of our candidate pathways were utilized as input parameters. The logistic regression analysis was performed using *glm*[55] function, and LOOCV analysis was prepared using *cv.glm* function from *boot* package in R.

To evaluate non-random predictive ability of the defined 5 candidate pathways, we used random model prepared by selecting 5 biological pathways at random, thereby we investigated the statistical significance of our finding by comparing the ability of the

candidate pathways to predict tamoxifen response to random equally-sized pathways. In details, 5 pathways were randomly chosen 1000 times from a total of 833 molecular pathways produced from Reactome, BioCarta and KEGG databases, and subsequently Kaplan-Meier survival analysis was utilized to evaluate the ability of random selected 5 pathways to predict therapeutic response. The empirical p-value for the significance was estimated as the number of times predictive ability of 5 random pathways reached or exceeded performance of our candidate 5 pathways.

2.7 Comparative analysis to other commonly utilized approaches

To assess the superiority of our approach over other commonly used techniques, we compared its performance to *(i)* extreme-responder analysis;[27] *(ii)* SVM;[30] and *(iii)* PRES random forest.[31] In each case, we utilized Training cohort for model training and Test cohort 1 for model validation. We compared groups of patients with poor and favorable response to tamoxifen in the Training cohort by selecting: patients that experienced events within 1 year of tamoxifen administration (i.e., *non-responders*, $n = 4$); and patients that did not experience any relapses for more than 9 years (i.e., *responders*, $n = 4$) to define a differential expression signature of tamoxifen response (i.e., through two-sample two-tailed Welch t-test[59] through *t.test* function in R). For Epsi et al method, we then subjected the differential expression signature to pathway enrichment analysis, where this signature was used as a reference and groups of genes from each pathway was used as a query gene set, and treated most significant pathways as candidate pathway markers. For SVM and PRES random forest, we subjected the differential expression signature (i.e., based on the proposed significance level) to the model training using Training cohort. The

SVM analysis was performed using *svm* function from *e1071* package, and PRES random forest analysis was prepared using *train* function from *caret* package in R. Predictive ability of the identified predictions was evaluated using Cox proportional hazards model through *survival* and *survminer* packages in R.

2.8 Pathway activity read-outs

To determine read-out genes of pathway's activity, we examined genes inside each molecular pathway, which were *(i)* changed on expression levels (i.e., leading edge genes from the single sample pathway enrichment output); *(ii)* correlated with pathway activity outputs (i.e., correlation analysis between a leading edge gene and NESs across all patients) through Spearman correlation using *cor.test* function in R ; and *(iii)* combined with tamoxifen response through univariable Cox proportional hazards model using adjusted hazard p-value (i.e., through *coxph* function). Finally, the adjusted hazard p-values were associated with Spearman correlation p-values to select the candidate genes.

2.9 Statistical analysis

Statistical analysis was performed using R studio version 3.5.1 for statistical computing. For single-sample analysis, data were z-scored on individual gene level. For this, the mean and standard deviation was first estimated for each gene across all samples in the dataset. Subsequently, z-score for each gene was defined as the difference between its own intensity value and the mean of that gene across the samples and divided by the standard deviation for that gene. The ranked list of z-scores for each gene in a sample then

defined single-sample signature. Pathway activity levels were estimated as Normalized Enrichment Scored (NESs) from the Gene Set Enrichment Analysis (GSEA), where NESs and p-values were estimated using 1,000 gene permutations. Cox proportional hazards model was utilized to associate pathway activity levels with treatment-related relapse-free survival (tRFS). When adjusting for common covariates multivariable Cox proportional hazards model was utilized, where its significance was reported using hazards ratio, hazards p-value, and Wald test. Kaplan-Meier survival analysis was utilized to estimate difference in treatment-related survival between two groups of patients, with log-rank p-value used to estimate significance.

2.10 Data availability

Data utilized for Training and Testing and their clinical characteristics are freely available from GEO repository GSE6532.

Chapter III

RESULTS

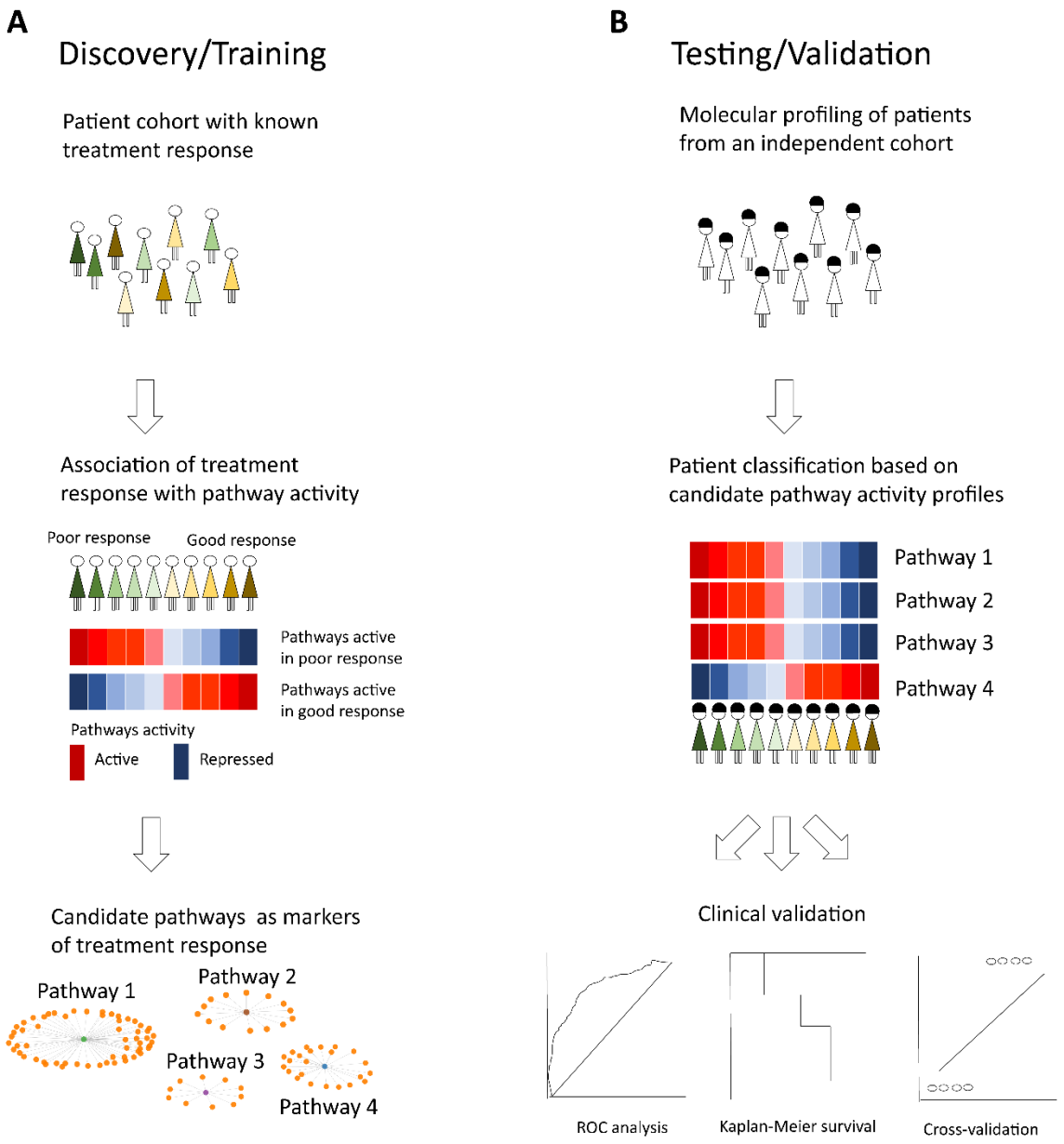
3.1 Overview

We present a genome-wide pathway-centric computational analysis to identify molecular pathways predictive of risk of resistance to tamoxifen in ER+ breast cancer patients. Our approach has the following steps:

Training phase (Figure 2A): *(i)* activity levels of biological pathways is estimated in each ER+ breast cancer patient (across a wide spectrum of responses, present in a clinical setting) that received adjuvant tamoxifen (Figure 1, Table 2); *(ii)* these pathway activity levels are then associated with tamoxifen treatment response across all patients, adjusted for common covariates;

Testing phase (Figure 2B): *(iii)* pathways that are significantly associated with the risk of tamoxifen resistance are then subjected to clinical validation analysis in independent patient cohorts (Figure 1, Table 2), for their ability to predict tamoxifen resistance for new incoming patients; *(iv)* finally, ability of the candidate pathways to predict the risk of tamoxifen resistance is compared to other known gene signatures of resistance and overall disease aggressiveness, alongside comparison to other methods.

Figure 2.



Schematic representation of the pathway-centric approach. (A) Training phase: identification of molecular pathways of tamoxifen resistance. (B) Testing phase: clinical validation of identified candidate pathways and multi-modal prediction evaluation.

3.2 Training phase: identifying molecular pathways that govern primary tamoxifen resistance

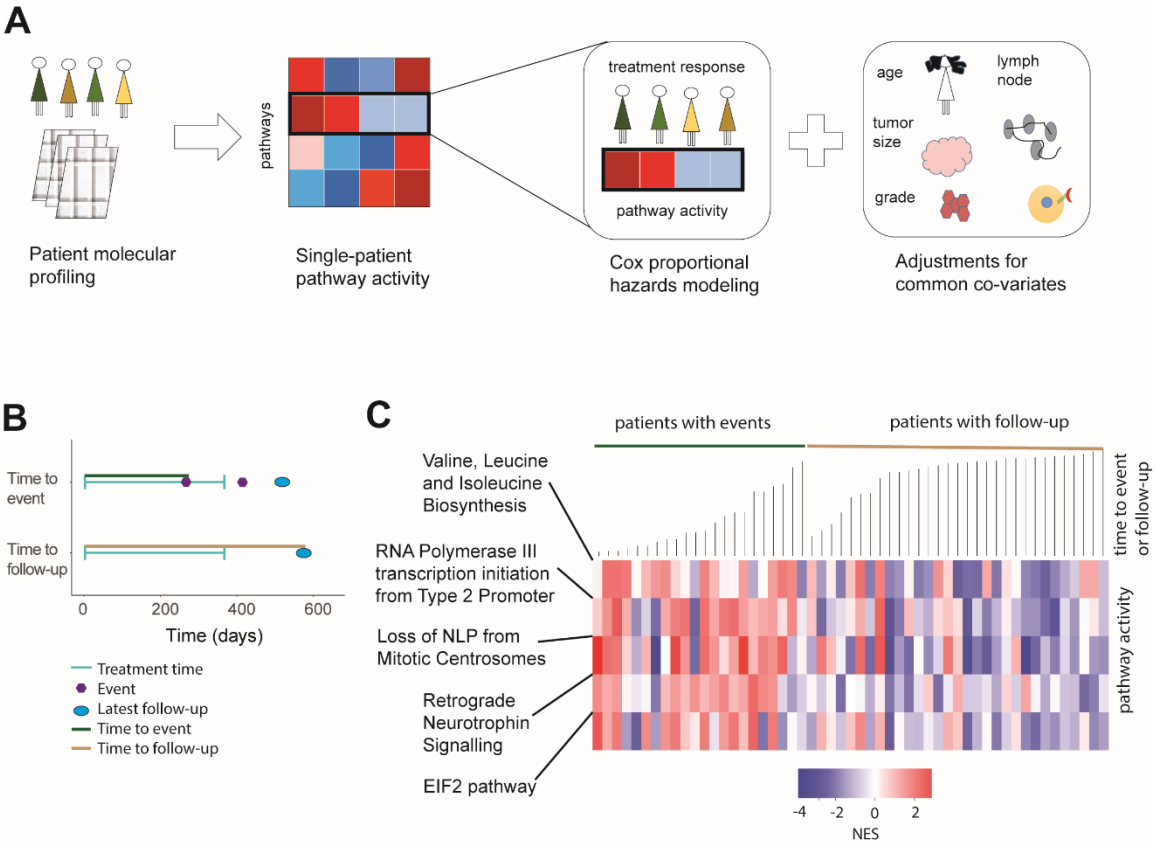
To accurately define therapeutic response to tamoxifen in ER+ breast cancer patients, we carefully selected gene expression profiles for the Training cohort (Loi et al.,[29] KIT-GSE6532) of primary ER+ breast tumors collected through surgery, not subjected to any neoadjuvant (i.e., prior to sample collection) treatment, and administered adjuvant (i.e., post-operative) 5-year long tamoxifen administration, with available clinical follow-up data (n = 57) (Figure 1, Table 2).

To avoid inconsistencies in BC classification, we subjected patient profiles of the Training cohort to a 50-gene Prediction Analysis of Microarrays panel[36] (PAM50) classification. PAM50 classification categorized BC patients from the Training cohort into the five intrinsic molecular subtypes: luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-enriched, triple-negative/basal-like, and normal-like, known to differ in their clinical outcomes[60,61] and therapy choice.[62] ER+ BC, which is the phenotype of interest in our study, is contained within the luminal A and luminal B subtypes and is excluded from HER2-enriched, triple-negative/basal-like, and normal-like subtypes (Table 2). Out of 57 post-operative tamoxifen-treated patients, 4 patients were classified as HER2-enriched, basal-like, or normal-like, and thus were excluded from further analysis.

Our objective was to evaluate tamoxifen response across all 53 patient samples (on the individual-patient level) and associate them with changes in biological pathway activities (Figure 3A). In order to be able to evaluate each patient sample individually, we scaled (i.e., z-scored, see Methods) gene expression profiles on individual gene levels so that each gene had mean 0 and standard deviation 1 over all samples in the Training

cohort.[39] The list of genes ranked by their z-scores in each sample then defined an individual-patient signature. We then utilized each individual-patient signature to evaluate activity levels of biological pathways using single-sample Gene Set Enrichment Analysis (ssGSEA),[44,45] where pathways were obtained from Reactome,[41] BioCarta[42] and KEGG[43] databases, corresponding to 833 pathways. For this analysis, each patient signature was used as a reference and each pathway as a query gene set. Activity levels of biological pathways were defined by their enrichment in each patient signature, mathematically represented by the Normalized Enrichment Scores (NES) from the GSEA analysis, where positive NES corresponds to enrichment in the over-expressed part of the signature and negative NES corresponds to enrichment in the under-expressed part of the signature (Figure 2A, Figure 3A).

Figure 3.



Training phase: pathway-centric approach identifies five biological pathways that govern tamoxifen response. (A) Schematic representation of the Testing phase of our approach: (*left*) patient molecular profiles are collected and analyzed; (*middle*) pathway activities are estimated in each patient using single-patient pathway enrichment analysis; (*right*) pathway activities are associated with response to tamoxifen using Cox proportional hazards modeling and are adjusted to common covariates, including age, tumor grade, tumor size (> 2 cm vs ≤ 2 cm), lymph node status, and PR status. (B) Graphical illustration of tamoxifen-related treatment response or follow-up. Time to event (top): time interval between tamoxifen administration and earliest relapse is indicated by green line. Time to follow-up (bottom): time interval between tamoxifen administration and latest follow-up date is indicated by brown line (no tamoxifen-related events observed). (C) Heatmap representation of the pathway activity levels (i.e., NES) and their association with time to tamoxifen-related relapse or follow-up, in the Training cohort. Green line marks the group of patients with tamoxifen-related relapse, sorted from the shortest to the longest time to relapse. Brown line marks the group of patients with follow-up and without disease relapse until the latest follow-up, sorted from the shortest to longest time to follow-up.

Next essential step in our analysis was to associate changes in pathway activity levels to tamoxifen treatment response. In general, we defined treatment-related relapse free survival (tRFS) as the interval between tamoxifen administration (which occurred immediately after surgery) and the earliest relapse (defined as local, regional, or distant metastasis) or the latest follow-up (these patients did not develop an event until their latest follow-up). When a patient had a relapse during or after the therapy administration, time to therapy related relapse was defined from therapy start to the earliest relapse (Figure 3B, top schematics, green line). When a patient never experienced a relapse, therapy-related relapse-free survival was measured from therapy start to the latest follow-up (Figure 3B, bottom schematics, brown line). In this dataset, 41.5% of patients experienced tamoxifen-related events (i.e., relapse), making it ideally suited for Training purposes.

To estimate association between the activity levels of the biological pathways and tRFS across a wide spectrum of tamoxifen response (taking into account a heterogeneity of response to tamoxifen, present in a clinical setting), we utilized Cox proportional hazards model,[46] ideally suited when time to event or follow-up is available. The Cox proportional hazards model was estimated between each pathway activity level (i.e., NESs, independent/predictor variable) and tamoxifen tRFS (i.e., dependent/response variable) across all 53 patients in the Training cohort. Furthermore, to account for the effect of other factors, this analysis was adjusted for commonly utilized covariates, as suggested in,[63] such as age, tumor grade, tumor size (> 2 cm vs ≤ 2 cm), lymph node status, and PR status (note that decreased PR levels are associated with increased HER2 signaling[16]) (Figure 3A). Such analysis identified five molecular pathways (Figure 3C, Table 3), most significantly associated with response to tamoxifen (hazard p-value ≤ 0.00075, Figure 4A-B, see

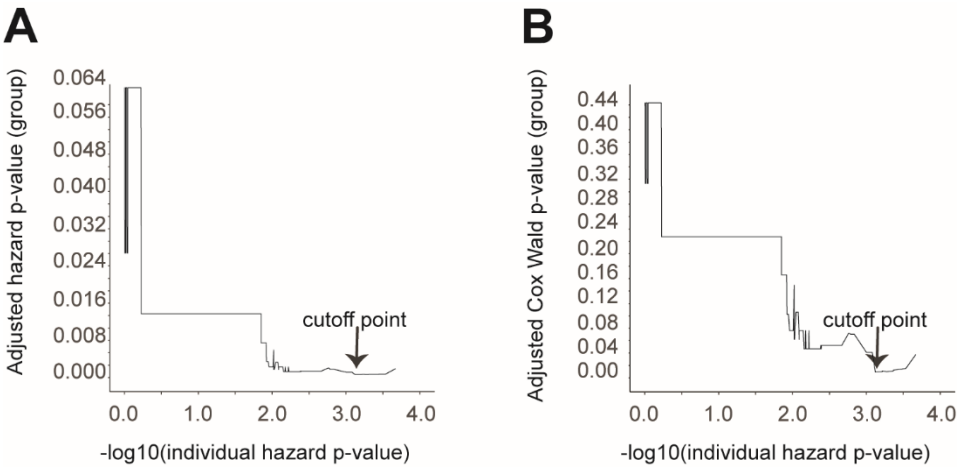
Methods), including Retrograde Neurotrophin Signalling, Loss of NLP from Mitotic Centrosomes, RNA Polymerase III Transcription Initiation from Type 2 Promoter, EIF2 pathway, and Valine Leucine and Isoleucine Biosynthesis, adjusting for parent-child relationships inherent in pathway databases (see Methods, Figure 5).

Table 3.

Five molecular candidate pathways and their corresponding significance levels.

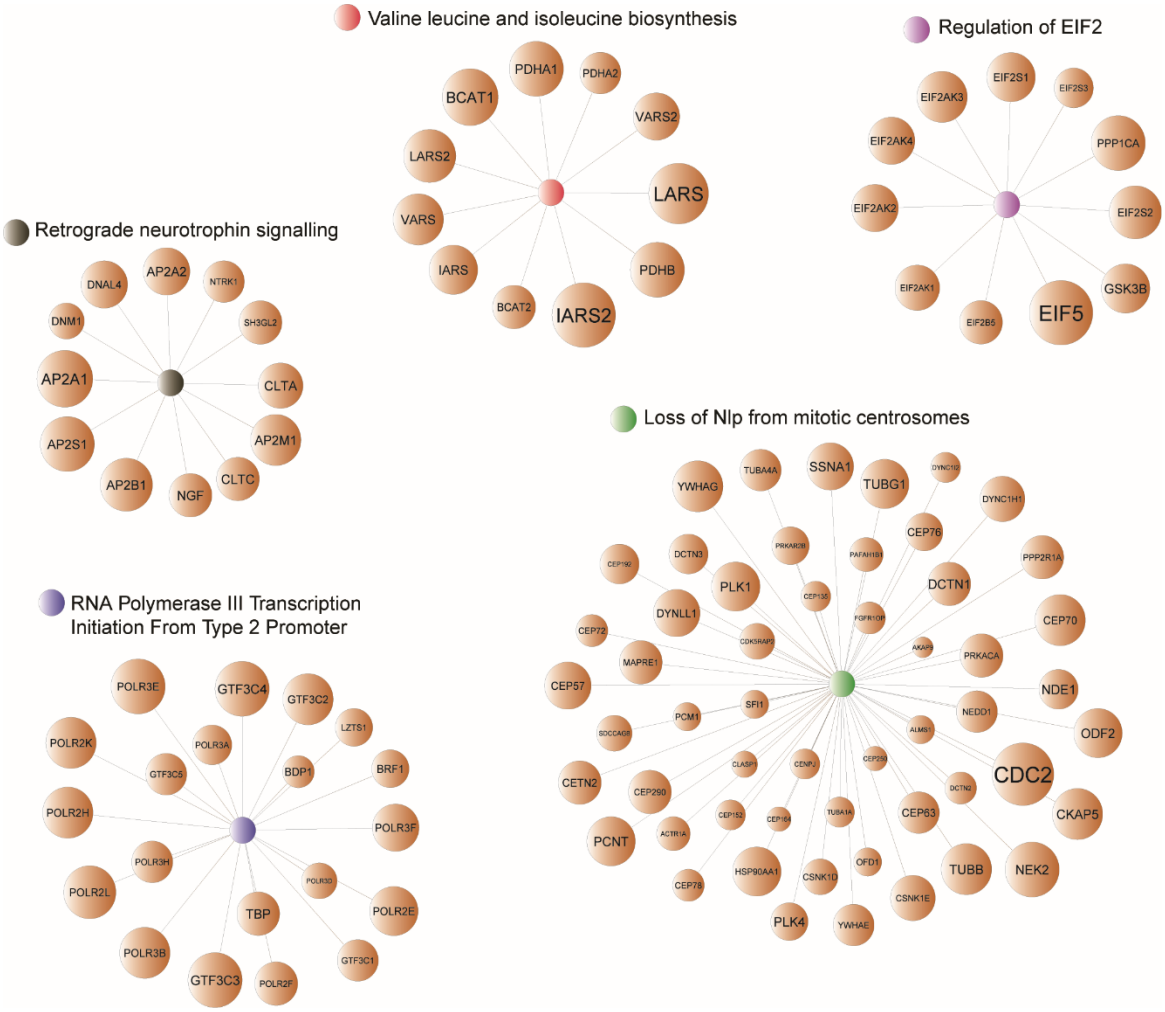
Pathway Names	Adjusted Hazard ratio (95%CI)	Adjusted Hazard p-value
REACTOME: RETROGRADE NEUROTROPHIN SIGNALLING	2.31(1.48-3.60)	0.00021
REACTOME: LOSS OF NLP FROM MITOTIC CENTROSOMES	1.73(1.28-2.33)	0.00029
REACTOME: RNA POLYMERASE III TRANSCRIPTION INITIATION FROM TYPE 2 PROMOTER	1.97(1.33-2.91)	0.0006
BIOCARTA: EIF2 PATHWAY	1.84(1.30-2.59)	0.00053
KEGG: VALINE LEUCINE AND ISOLEUCINE BIOSYNTHESIS	1.78(1.27- 2.50)	0.00075
Note: CI: confidence intervals.		

Figure 4.



Comprehensive threshold analysis identifies pathway significance level. Threshold analysis in the Training cohort utilizing Cox proportional hazards model on the group of pathways (starting from the most significant pathways and adding the next most significant pathway, one at a time). Cutoff point was determined as a point on the graph when adding any additional pathway would not improve the model significance. Adjusted hazard p-value (A) and adjusted Cox Wald p-value (B) are used as threshold-deciding criteria.

Figure 5.



Graphical representation of the five candidate pathways and their significantly contributing genes. Network-based representation of the five candidate pathways. Selected genes shown (brown nodes) correspond to genes that contribute to significant enrichment of each pathway in the patient single-sample signatures. Node sizes represent number of times each gene appears in the leading edge in the single-sample pathway enrichment analysis (i.e., indicating significant changes in activity of this pathway across Training cohort).

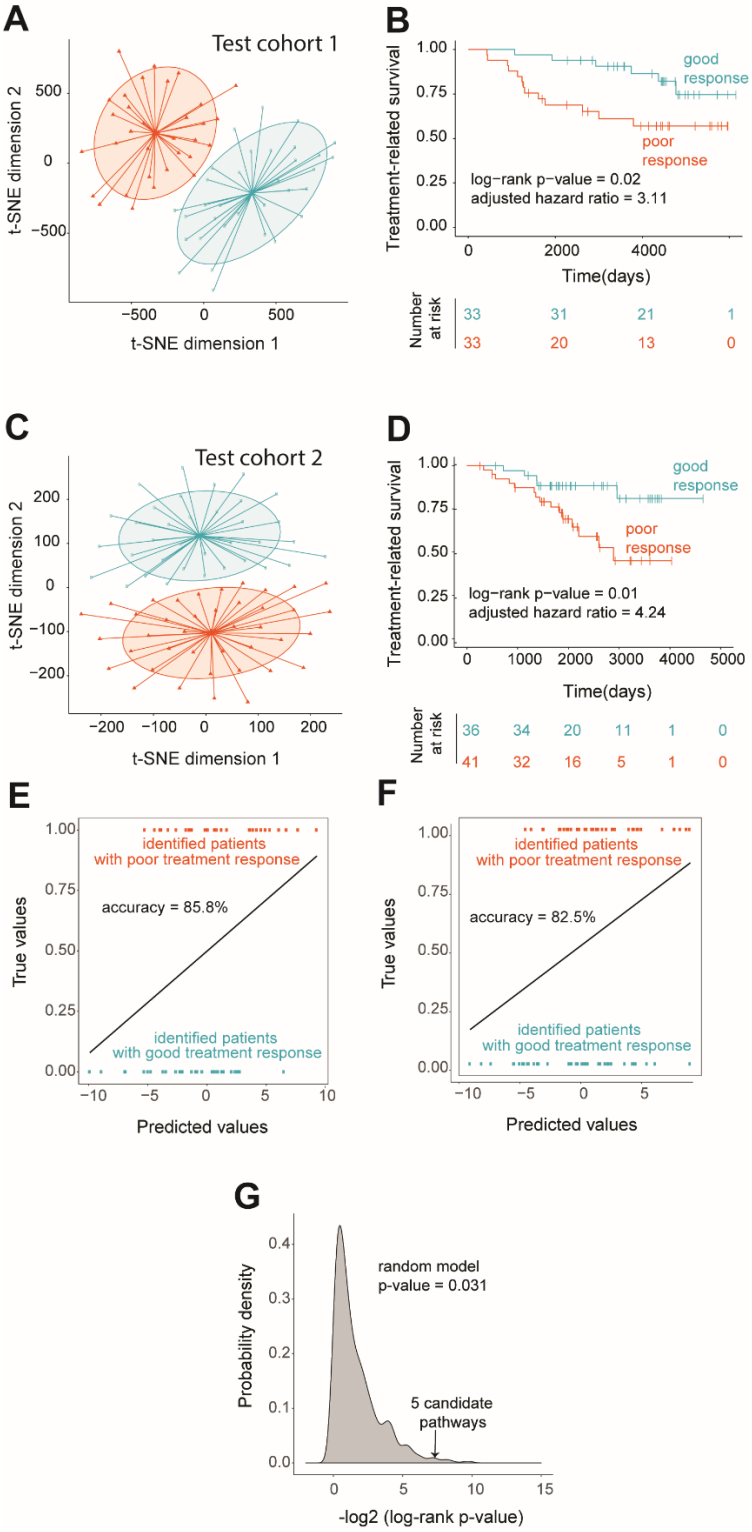
3.3 Testing phase: clinical validation in independent patient cohorts

The next essential step in our analysis was to evaluate the ability of five candidate pathways to predict treatment response to tamoxifen in independent non-overlapping clinical cohorts. For this, we utilized two patient cohorts for testing/validation purposes: *(i)* Test cohort 1[29] (GUYT-GSE6532, n = 70) of primary breast tumors obtained at surgery, from patients that did not receive any neoadjuvant treatment and received only adjuvant tamoxifen, with 28.78% of patients having tamoxifen-related events (Table 2); and *(ii)* Test cohort 2[29] (OXFT-GSE6532, n = 77) of primary breast tumors obtained at surgery, from patients that did not receive any neoadjuvant treatment and received only adjuvant tamoxifen, with 25.97% of patients with tamoxifen-related events. Both Test cohorts had clinical characteristics, neoadjuvant, and adjuvant conditions comparable to the Training cohort (Table 2). Similar to the analysis done on the Training cohort, we performed PAM50 classification on the two Test cohorts, eliminating 4 patients from Test cohort 1 and keeping all patients for Test cohort 2.

Our main objective was to investigate if activity levels of the five candidate pathways could predict risk of resistance to tamoxifen in two independent Test cohorts. For this, we estimated activity levels for five candidate pathways in each patient in the Test cohorts (similarly to Training cohorts, see Methods) and subjected patients to t-distributed Stochastic Neighbor Embedding (t-SNE) clustering[48] as suggested in[49] for investigation of samples relationships. T-SNE analysis, which displays five-dimensional dataset in a two-dimensional space, stratified patients into two groups based on their pathway activity levels. The low-dimensional output (i.e., 2-dimensional) of t-SNE were then subjected to the k-means clustering[51] to correctly assign group membership (Figure 6A for Test cohort

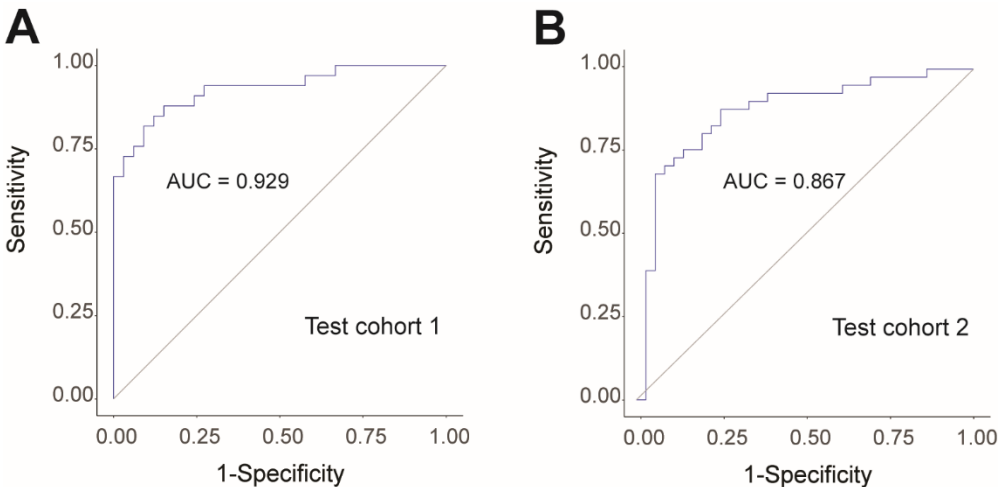
1 and Figure 6C for Test cohort 2) one group with increased pathways' activities (orange) and one group with decreased pathways' activities (turquoise), mimicking the relationship that was observed in the Training cohort (Figure 3C). We confirmed the strength of group separation through Receiver Operating Characteristic (ROC) analysis[53] using multiple logistic regression model (Figure 7A-B), where normalized enrichment scores of 5 pathways were used as input parameters (i.e., independent/predictor variables) and selected patient groups were utilized as a dependent/response variable. The efficiency of ROC analysis was estimated using area under the curve (AUC),[54] where AUC of 0.5 denotes a random predictor and AUC score of 1 denotes a perfect predictor (i.e., full separation of the patient groups). This analysis confirmed that the activity levels of the five candidate pathways can be effectively used for classifying patients into distinct groups (Test cohort 1, AUC = 0.929; Test cohort 2, AUC = 0.867).

Figure 6.


A
Test cohort 1
t-SNE dimension 2
t-SNE dimension 1
B
Treatment-related survival
good response
poor response
log−rank p−value = 0.02
adjusted hazard ratio = 3.11
Time(days)
Number at risk
33 31 21 1
33 20 13 0
C
Test cohort 2
t-SNE dimension 2
t-SNE dimension 1
D
Treatment-related survival
good response
poor response
log−rank p−value = 0.01
adjusted hazard ratio = 4.24
Time(days)
Number at risk
36 34 20 11 1 0
41 32 16 5 1 0
E
True values
identified patients with poor treatment response
accuracy = 85.8%
identified patients with good treatment response
Predicted values
F
True values
identified patients with poor treatment response
accuracy = 82.5%
identified patients with good treatment response
Predicted values
G
Probability density
random model p-value = 0.031
5 candidate pathways
-log2 (log-rank p-value)

The five candidate pathways predict patients at risk of tamoxifen resistance in independent patient cohorts. (A, C) T-SNE and subsequent k-means clustering of Test cohort 1 (A) and Test cohort 2 (C) based on activity levels of the five candidate pathways demonstrates patient separation into two groups: orange group (with overall increased activity levels of the five candidate pathways) and turquoise group (with overall decreased activity levels of the five candidate pathways). (B, D) Kaplan-Meier treatment-related survival analysis comparing two patient groups in Test cohort 1 (B) and in Test cohort 2 (D). Log-rank p-values and adjusted hazard ratios are indicated. (E, F) Leave-one-out cross-validation (LOOCV) correctly identified patients with poor response to tamoxifen (orange) and patients with favorable response to tamoxifen (turquoise) in Test cohort 1 (E) and Test cohort 2 (F). Accuracy values (%) are indicated. (G) Random model to assess the ability of the 5 candidate pathways selected at random to differentiate samples into groups with various tamoxifen response. The significance of the predictive ability of our defined 5 molecular pathways is shown by the distributions of log-rank p-values from the random model.

Figure 7.



ROC analysis demonstrated significant separation of patient groups based on activity levels of the five candidate pathways. ROC analysis to show significance of the separation between patient groups in Figure 6 A, C. Area under the curve (AUC) is reported.

To assess if these patient groups significantly differ in their tamoxifen response, we analyzed therapy-related relapse-free survivals between the groups using Kaplan-Meier survival analysis[57] and Cox proportional hazards model,[46] which demonstrated that the

identified patient groups had a significant difference in their response to tamoxifen (Test cohort 1, log-rank p-value = 0.02, Figure 6B; Test cohort 2, log-rank p-value = 0.01, Figure 6D). We have also adjusted these analyses for common covariates[63] (i.e., age, tumor grade, tumor size, lymph node status, and PR status), demonstrating that these covariates did not significantly impact the predictive ability of our findings (Test cohort 1, adjusted hazard ratio = 3.11, adjusted hazard p-value = 0.044, 95% confidence interval CI: = 1.03-9.396, Figure 6B; Test cohort 2, adjusted hazard ratio = 4.24, adjusted hazard p-value = 0.012, CI: 1.3708- 13.120, Figure 6D).

Further, we evaluated predictive accuracy of our model in the two test cohorts using Leave-one-out cross-validation (LOOCV), which simulates a situation when a new incoming patient needs to be evaluated for her risks of developing resistance to tamoxifen. In particular, in LOOCV, one patient is "removed", and the model is trained on the remaining patients, followed by the prediction of risk of resistance for the removed patient. The process is repeated for each patient. Using this analysis, we demonstrated the accurate performance of our model in predicting poor and favorable tamoxifen response for new incoming patients (Test cohort 1, accuracy for LOOCV = 85.8%, Figure 6E; Test cohort 2, accuracy for LOOCV = 82.5%, Figure 6F). Finally, to evaluate if any set of five pathways selected at random could classify patients based on tamoxifen response, we have performed random model analysis, where selected five pathways at random 1,000 times and evaluated their predictive ability using Kaplan-Meier survival analysis, as above. The empirical p-value for the random model was estimated as the number of times log-rank p-values of five pathways chosen at random reached or exceeded the output of our 5 candidate pathways, which confirmed significant non-randomness of our candidate pathways predictive ability

(Test cohort 1, random model p-value = 0.031, Figure 6G). Taken together, these findings indicate that the five-candidate pathway signature could successfully predict patients at risk of tamoxifen resistance in independent patient cohorts.

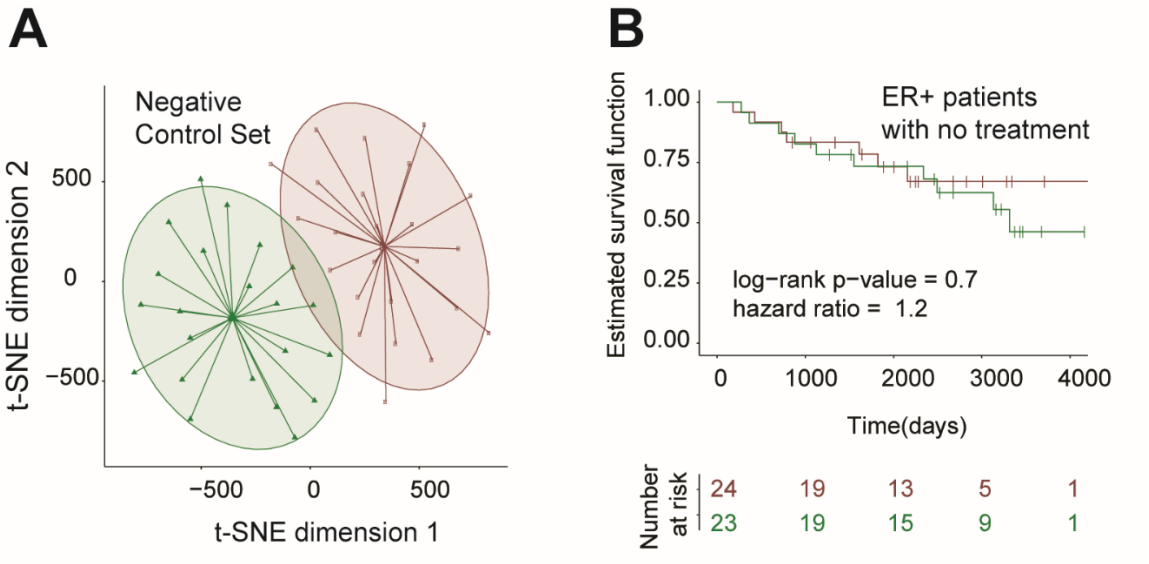
3.4 Comprehensive comparison of tamoxifen response and overall disease aggressiveness

A fundamental question in studying therapeutic response lies in its comparison to and differentiation from overall disease aggressiveness. Our comprehensive investigation of this question was four-fold: *(i)* we identified pathways implicated in disease aggressiveness and compared their overlap with the candidate five pathways of tamoxifen response; *(ii)* we evaluated if the five candidate pathways can predict breast cancer aggressiveness in an independent (negative control) cohort; *(iii)* we evaluated the ability of the five candidate pathways to predict tamoxifen response based on PR status (known indicator of breast cancer aggressiveness), age categories (as patients aged 50 years or older shows poorer relative survival rates than younger patients [64]) and luminal subtypes (as luminal B type have a poorer prognosis than luminal A type [65]); and *(iv)* we evaluated if known published signatures of disease aggressiveness could predict response to tamoxifen.

First, to examine if our 5 candidate pathways overlap with pathways implicated in disease aggressiveness, we developed treatment-free prognostic pathway signature using a patient cohort that received surgery only (KIU-GSE6532, n = 51, negative control cohort).[29] Out of 51 surgery-treated patients, 4 patients were removed, based on the PAM50 classification. We further applied our single-sample pathway-based discovery approach (as in the Training phase) and associated them to the RFS, which identified 3 pathways of

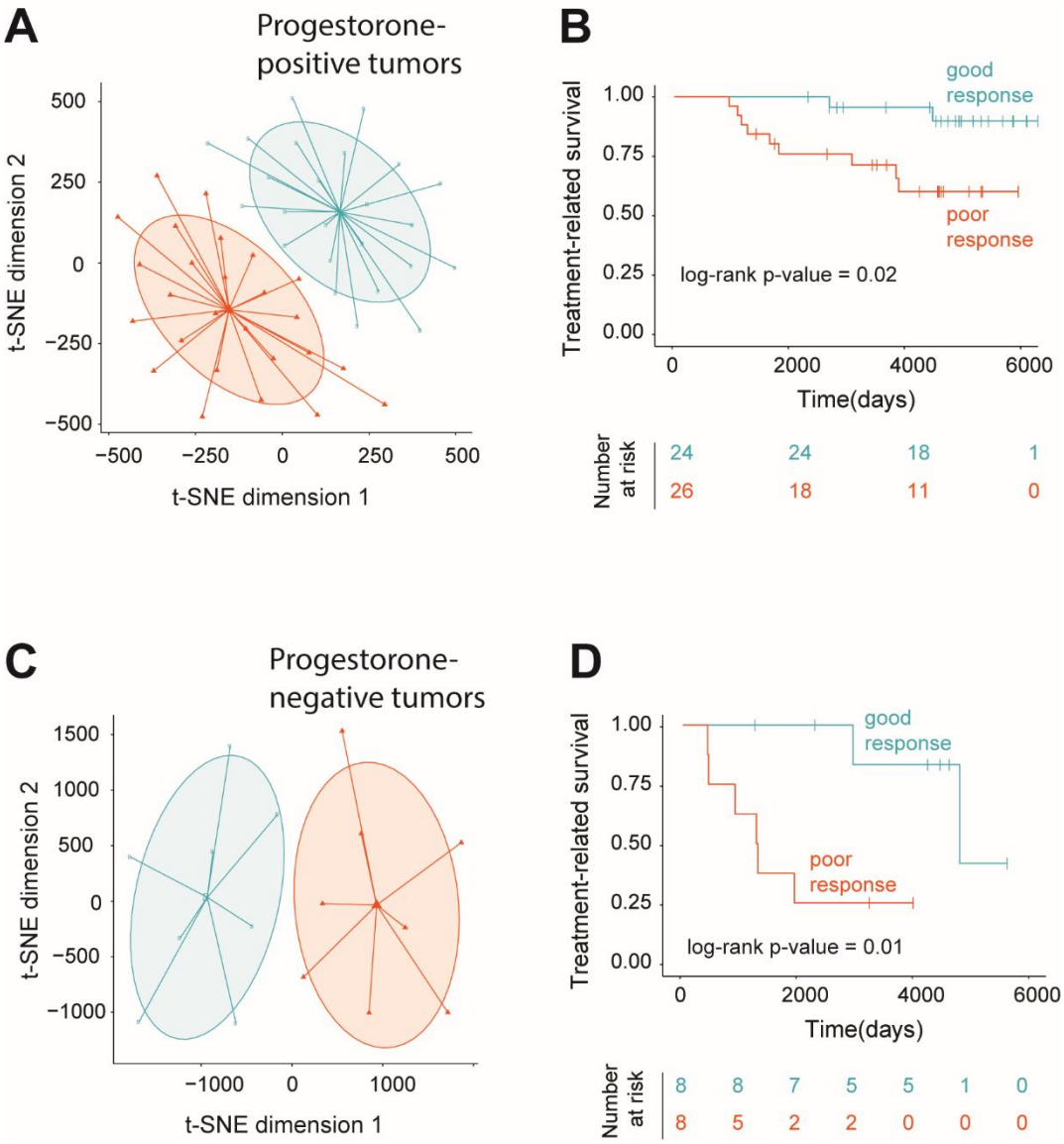
aggressiveness (see Methods) that showed no overlap with the five candidate pathways, signifying that none of our candidates are involved in cancer severity and are indeed specific to tamoxifen response.

Figure 8.



Five candidate pathways do not predict overall disease aggressiveness. (A) T-SNE and subsequent k-means clustering based on the activity levels of the five candidate pathways in the negative control cohort. (B) Kaplan-Meier survival analysis on negative control cohort confirms that the five candidate pathways do not predict disease aggressiveness. Log-rank p-value and hazard ratio are indicated.

Figure 9.



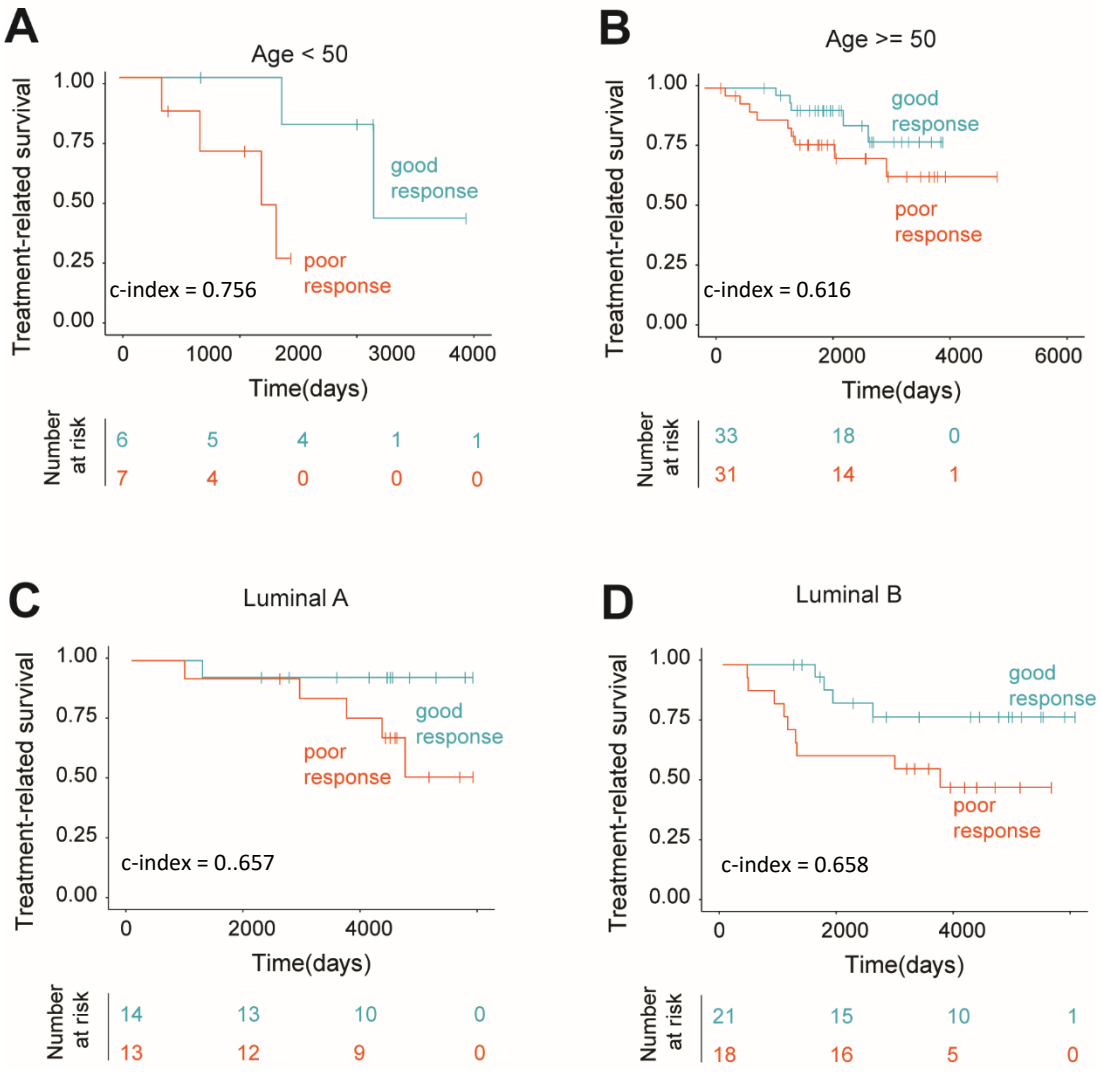
Stratified analysis demonstrates that predictive ability of the five candidate pathways is not dependent on the PR status. Patients in Test cohort 1 were stratified based on their progesterone PR status: Progesterone-positive (A, B) and Progesterone-negative (C, D). T-SNE with subsequent k-means clustering on PR+ (A) and PR- (C) patient subgroups. Kaplan-Meier survival analysis for PR+ (B) and PR- (D) patient groups. Log-rank p-values are indicated.

Secondly, we evaluated if the five candidate pathways could separate patients based on overall disease aggressiveness. For this, we evaluated predictive ability of the five candidate pathways on the BC patient cohort that did not receive any treatment after surgery (negative control cohort, as above). We subjected the dataset to the single-sample pathway enrichment analysis (for the five candidate pathways, similarly to Test cohorts analysis). T-SNE clustering (Figure 8A) and subsequent Kaplan-Meier survival analysis (Figure 8B) on this cohort demonstrated that the five pathways do not separate patients base on their disease aggressiveness (hazard ratio = 1.2, log-rank p-value = 0.7, RFS was considered as a clinical endpoint), but rather specific for tamoxifen response. We have also examined the effect of covariates (i.e., age, tumor grade, tumor size, and PR status), on disease progression in this setting and demonstrated that our candidate pathways remain insignificant, with tumor size significantly contributing to the disease progression (adjusted hazard p-value = 0.0307).

Third, given that the PR receptor status (which also reflects HER2 signaling) is a known indicator of breast cancer aggressiveness, we performed a stratified Kaplan-Meier analysis on Test cohort 1 (for which this information was available). For this, we divided Test cohort 1 into two groups: one with PR-positive status and one with PR-negative status. We then subjected both cohorts separately to t-SNE clustering, which have demonstrated that the five candidate pathways separated each cohort into patient sub-groups with high and low levels of pathway activities (Figure 9A for patients with PR-positive tumors and Figure 9C for patients with PR-negative tumors). Subsequent Kaplan-Meier survival analysis (Figure 9B and Figure 9D, respectively) showed that these patient-subgroups significantly differ in their response to treatment (patients with PR-positive tumors, log-

rank p-value = 0.02, Figure 9B; patients with PR-negative tumors, log-rank p-value = 0.01, Figure 9D), demonstrating that our five candidate pathways are able to predict patients at risk of tamoxifen resistance regardless of the PR-status.

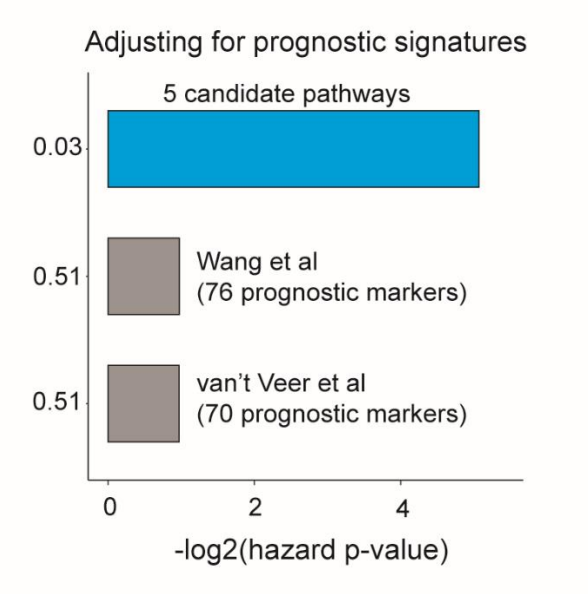
Figure 10.



Stratified analysis demonstrates that predictive ability of the five candidate pathways does not depend on the age groups and luminal subtypes. Stratified Kaplan-Meier survival analysis was conducted using different age groups in Test cohort 2 (A, B) and luminal subtypes in Test cohort 1 (C, D).

In fact, different age categories and luminal subtypes are also known indicators of worse prognosis in breast cancer. Thus, we conducted stratified Kaplan-Meier analysis (as above); where we stratified Test cohort 2 into patient groups based on age (< 50 years and >=50 years), and Test cohort 1 into patient groups based on luminal subtypes (luminal A and luminal B). Kaplan-Meier survival analysis showed clear separation between patient groups (Figure 10 A-D). Our analysis demonstrated that the predictive ability of these candidate pathways does not depend on known characteristics of breast cancer aggressiveness.

Figure 11.



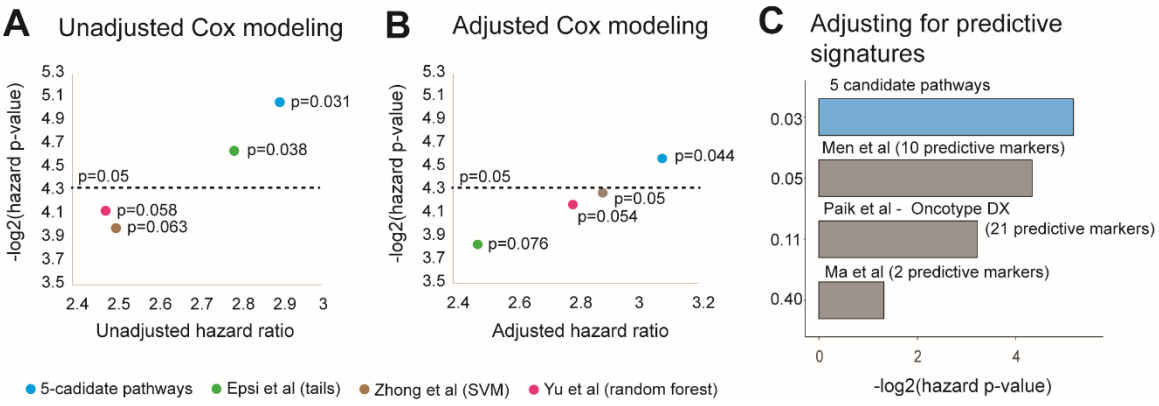
Five candidate pathways are not affected by overall disease aggressiveness. Multivariable Cox proportional hazards model representing analysis for five candidate pathways adjusted for various prognostic signatures in breast cancer, including Wang et al. (76 prognostic markers, with 57 present on U133 Plus 2.0) and van't Veer et al. *(*70 prognostic markers, with 53 present on U133 Plus 2.0). Adjusted hazard p-values are reported.

Finally, to demonstrate that the predictive ability of the five candidate pathways is not affected by other known markers of disease aggressiveness, we investigated if commonly known gene-based prognostic signatures can predict tamoxifen response or affect predictive ability of the candidate five pathways. For this, we gathered several known signatures of overall BC aggressiveness (i.e., prognostic signatures), including Wang et al. signature[66] (76 prognostic markers, with 57 present on U133 Plus 2.0) and van't Veer et al. signature[67] (70 prognostic markers, with 53 present on U133 Plus 2.0) and subjected them to adjusted multivariable Cox proportional hazards model, alongside the five candidate pathway signature, in the Test cohort 1. This analysis confirmed that the prognostic signatures were not predictive of tamoxifen response and did not impact predictive ability of the five candidate pathways (adjusted hazard p-value = 0.03, Figure 11). Taken together, these findings indicate that our five-pathway signature of tamoxifen response is not indicative of overall breast cancer aggressiveness and is indeed specific to response to tamoxifen.

3.5 Comparative analysis to commonly utilized methods and known signatures of tamoxifen response

To evaluate predictive advantages of the five candidate pathways, we took a comprehensive approach and first *(i)* compared the predictive ability of the five candidate pathways to predictions from other commonly used methods, including approaches based on extreme-responder analysis (i.e., tails of the distribution), support vector machine (SVM), and random forest; and second *(ii)* assessed if the predictive ability of the five candidate pathways outperforms other known signatures of tamoxifen response.

Figure 12.



Predictive ability of the five candidate pathways outperforms markers from other methods and known signatures of tamoxifen response. (A, B) Comparison of the predictive ability of the five candidate pathways (blue) to the candidate identified by other approaches, including Epsi et al. extreme-responder analysis (green), Zhong et al. SVM-based method (brown) and Yu et al. PRES random forest-based method (pink), through unadjusted (A) and adjusted for common covariates (B) Cox proportional hazards model. P-values for unadjusted and adjusted hazard ratios are indicated. (C) Multivariable Cox proportional hazards model representing analysis for the five candidate pathways adjusted for different predictive signatures of tamoxifen response, including Men et al. (10 predictive markers, with 9 present on U133 Plus 2.0), Paik et al. (Oncotype DX, 21 predictive markers), and Ma et al. (2 predictive markers). Adjusted hazard p-values are indicated.

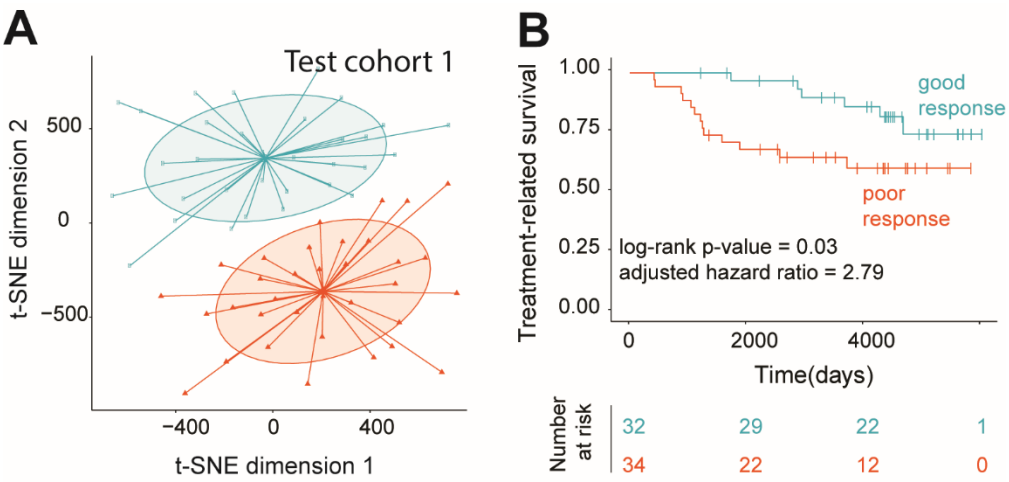
First, we compared predictive ability of the five candidate pathways to predictions from other commonly utilized methods, such as *(i)* Epsi et al.[27] method, which utilized extreme-responder analysis, using tails of the treatment response distribution to define a treatment response signature; *(ii)* Zhong et al.[30] method , which used Support Vector Machine approach as a base; and *(iii)* Yu et al.[31] method, also referred to as Personalized REgimen Selection (PRES), which used random forest approach as a base (see Methods). To assure that all methods are comparable to our pathway-centric method, we trained Epsi et al., Zhong et al., and Yu et al. methods on the Training cohort, with each producing a list of predictions (112 predictions for Epsi et al.; 5 predictions for Zhong et al.; and 3

predictions for Yu et al.). We then followed by validating these predictions on the Test cohort 1, similarly to our pathway-centric method. Such analysis demonstrated that the five candidate pathways outperform all three methods in their ability to predict the risk of tamoxifen treatment resistance (Figure 12A: five candidate pathways, hazard ratio = 2.91, hazard p-value = 0.031; Epsi et al., hazard ratio = 2.79, hazard p-value = 0.038; Zhong et al., hazard ratio = 2.53, hazard p-value = 0.063; Yu et al., hazard ratio = 2.48, hazard p-value = 0.058). Furthermore, we adjusted these analyses for the effect of common covariates (similarly to our original training phase), including age, tumor grade, tumor size, lymph node status and PR status and re-confirmed that the five candidate pathways retain their significant predictive ability and outperform the other methods (Figure 12B: five candidate pathways, adjusted hazard ratio = 3.11, adjusted hazard p-value = 0.044; Epsi et al., adjusted hazard ratio = 2.48, adjusted hazard p-value = 0.076; Zhong et al., adjusted hazard ratio = 2.96, adjusted hazard p-value = 0.05; Yu et al., adjusted hazard ratio = 2.81, adjusted hazard p-value = 0.054).

Finally, to confirm that the predictive ability of the five candidate pathways outperforms other known signatures in their ability to predict tamoxifen treatment response, we selected known signature of tamoxifen response (i.e., predictive signatures), such as *(i)* Men et al.[18] (10 predictive markers, with 9 present on U133 Plus 2.0); *(ii)* Paik et al.[19] (also now as Oncotype DX, 21 predictive markers); and *(iii)* Ma et al.[20] (2 predictive markers) (Figure 12C) and used them in adjusted multivariable Cox proportional hazards model, alongside the five candidate pathway signature, utilizing Test cohort 1, as above. This analysis demonstrated that the additional predictive signatures do not significantly affect the ability of the five candidate pathways to predict the risk of tamoxifen resistance

(Figure 12C, adjusted hazards p-value = 0.03). Taken together, these results demonstrate that the five-candidate pathway signature can be utilized to predict patients at risk of developing resistance to tamoxifen in a clinical setting and build a foundation for personalized therapeutic advice for patients with ER+ breast cancer.

Figure 13.



The five read-out genes predict patients at risk of tamoxifen resistance in independent patient cohort. (A) T-SNE and subsequent k-means clustering of Test cohort 1 based on gene expression of the five read-out genes demonstrates patient separation into two groups. (B) Kaplan-Meier treatment-related survival analysis comparing two patient groups in Test cohort 1. Log-rank p-value and hazard ratio are indicated.

Table 4.

Five read-out genes and their corresponding significance levels.

Read-out genes	Spearman correlation p-value	Adjusted Hazard p-value
AP2S1	0.0000278	0.0185
CDC2	0.00000035	0.0063
GTF3C3	0.00000004	0.00133
EIF2AK3	0.00000542	0.019
LARS	0.00000013	0.019

3.6 Pathway activity read-outs

For this, we examined genes which could function as read-outs of activity levels of pathways that involved in treatment response. In details, we examined genes inside each molecular pathway, which were *(i)* changed on expression levels (i.e., from leading edge in pathway enrichment outputs); *(ii)* correlated with pathway activity outputs (i.e., NESs); and *(iii)* combined with tamoxifen response (see Methods, Table 4). This analysis defined 5 read-out genes (i.e., AP2S1, CDC2, GTF3C3, EIF2AK3, and LARS) that were significantly related to therapeutic response (Test cohort1: log-rank p-value = 0.03, hazard ratio = 3.76, Figure 13 A-B). In fact, these findings showed a high predictive ability in identifying patients at risk of resistance. We suggest that these 5 read-out genes can be utilized as biomarker of tamoxifen resistance and will be more useful in the clinic.

Chapter IV

DISCUSSION

In this study, we have demonstrated that a pathway-centric genome-side computational approach is able to uncover biological pathways, highly associated with risk of tamoxifen resistance in ER+ breast cancer patients. The important advantage of our approach is that it identifies a tightly connected group of genes - biological pathways - as opposed to individual (possibly distantly connected genes), thus *(i)* decreasing the chances of *experimental noise* present in biological experiments; *(ii)* improving our understanding of the *mechanisms implicated in therapeutic resistance*; and *(iii)* increasing the likelihood of identifying a *functionally relevant signature*, which could be utilized to study mechanisms of primary resistance and their potential therapeutic targeting. Furthermore, these biological pathways are highly associated with a wide spectrum of treatment responses (as opposed to selecting a limited category of patients for analysis), reflecting heterogeneity of response to tamoxifen present in a clinical setting. Even though this work is focused on identifying cases of resistance to tamoxifen, our method can be broadly applicable to other therapeutic interventions and cancer types.

Our computational analysis has identified five molecular pathways implicated in tamoxifen resistance, including *(i)* Retrograde Neurotrophin Signalling, *(ii)* Loss of NLP from Mitotic Centrosomes, *(iii)* RNA Polymerase III Transcription Initiation from Type 2 Promoter, *(iv)* EIF2 pathway, and *(v)* Valine Leucine and Isoleucine Biosynthesis. Interestingly, many of these pathways have been shown to be closely related to carcinogenic mechanisms and therapeutic response in various cancers. In particular, the

Retrograde Neurotrophin Signalling pathway is implicated in metabolic detoxification, mitosis, clathrin-mediated vesicles development, and enriched with bladder cancer predisposition loci.[68] One of the genes from this pathway, Neurotrophic tyrosine kinase receptor type 1 (NTRK1), is a recognized oncogene frequently altered in various tumor types[69] and its gene fusions have previously been identified in glioblastoma,[69] colon cancer,[70] papillary thyroid carcinoma,[71] and non-small cell lung cancers.[72] Clinical studies of tumor response to NTRK1 fusion-targeted therapy have indicated that this oncogene represents a treatment target in human cancer.[73]

Ninein-like protein (NLP) (i.e., also known as NINL) is a part of the Loss of NLP from the Mitotic Centrosomes pathway. The role of human centrosomal NLP expression in breast, lung, ovarian, head and neck cancers has been widely demonstrated.[74] The NLP gene amplification accounts for NLP overexpression in human breast and lung cancer cells.[74] The deregulated expression of NLP in cell models leads to mitotic spindle aberrations, spindle checkpoint defects, chromosomal missegregation, cytokinesis failure, stimulation of chromosomal instability, anchorage-independent growth, and cell malignant transformation.[74] Recently, it has been discovered that NLP co-localizes and interacts with BRCA1 at inter-phasic centrosome and thus the disruptions of BRCA1 function could affect NLP co-localization to centrosomes and induce the genomic instability.[75] Interestingly, it has been reported that the NLP overexpression may also cause breast cancer resistance to paclitaxel chemotherapy.[76] Furthermore, a positive correlation between expression of NLP and PLK1 (i.e., another gene implicated in the Loss of NLP from the Mitotic Centrosomes pathway) has recently been discovered, implicated in

chemoresistance, particularly to taxane agents[76] and tumor growth in general, in breast cancer and other cancer types.[76,77]

For the RNA Polymerase III Transcription Initiation from RNA polymerase II Promoter Sites, the global gene expression is increased in eukaryotic cells as RNA polymerase II transcribes protein-coding genes to yield mRNA, miRNA, snRNA, and snoRNA genes while RNA polymerase III transcribes the genes for 5S rRNA and tRNAs.[78,79] However, in yeast (S. cerevisiae), RNA polymerase III complex has been shown to act as heterochromatin barriers,[80] and any change in heterochromatin would be potentially very important for cancer development. A broad spectrum of cancer cell types has been observed to display a highly regulated and elevated level of RNA polymerase III transcript expression.[81,82]

In the Eukaryotic Initiation Factor 2 (eIF2) pathway, phosphorylation of eIF2α has been shown to play a significant role in maintaining normal cellular homeostasis and regulating cell growth,[83] with dysregulation of eIF2 signaling pathway stimulating the cancerous tumors transformation.[84] The overexpression of eIF2α has been observed in several cancers, such as gastrointestinal cancer[85] and non-Hodgkin's lymphomas[86] and has been proposed as a potential therapeutic target.[87]

Finally, in the Valine Leucine and Isoleucine Biosynthesis pathway, valine, leucine, and isoleucine are important branched-chain amino acids (BCAAs) for normal growth and development.[88] In the BCAA catabolism pathway, the first step is transamination, catalyzed by the branched chain amino acid transferase isozymes BCATs: a mitochondrial (BCATm) and a cytosolic (BCATc) isozyme.[89-91] Mitochondrial BCATm (*BCAT2*) expression can drive the development of pancreatic ductal adenocarcinoma under the

regulation of the mitochondrial malic enzyme 2.[92,93] Malic enzyme 2 and malic enzyme 3 are oxidative decarboxylation that stimulate malate to pyruvate and are considered important elements during mitochondrial reactive oxygen species homeostasis and NADPH production.[93] Studies have reported that incorporating the molecular and metabolomic examination of malic enzyme-deficient cells showed a decrease in NADPH generation and an increase in reactive oxygen species level. [93] These alterations catalyze AMPK which consequently functions to inhibit SREBP1 and thereby suppresses its target genes including the BCAT2. BCAT2 stimulates the transfer of the amino group of BCAAs to α-ketoglutarate and thus generates glutamate, which in turns enhance *de novo* nucleotide synthesis. Therefore, the deficiency of the mitochondrial malic enzyme, which causes a reduction in NADPH synthesis, plays a critical role in the therapeutic strategy for the treatment of patients with complex disease. [93]

Cytosolic BCATc (BCAT1) is overexpressed in glioblastoma,[94] nasopharyngeal carcinoma,[95] and cancers with elevated c-MYC.[95] It has been recommended to consider BCAT1 as a promising target for glioblastoma and nasopharyngeal carcinoma treatments.[94,95] We propose that the identified candidate pathways should be investigated for their potential use as isolated treatment targets or in combination with ER-targeting agents for ER+ breast cancer patients at risk of developing resistance to tamoxifen.

One of the limitations of our study is in the limited availability of the epigenomic profiles for our patient cohorts. In fact, DNA and histone methylation has been suggested to be responsible for inactivation of ER.[96] Thus, further examination of the role of epigenomic modulations and their interplay with transcriptomic changes is an invaluable

next step for in-depth understanding of molecular mechanisms implicated in hormone therapy resistance.

Furthermore, miRNAs (micro-RNAs) have received substantial attention for their role in regulating pathway functionality.[97] For example, miR-15a/miR-16's deletion or down-regulation contributes to dysregulated of cell cycle in chronic lymphocytic leukemia[98] and non-small cell lung cancer.[99] Even though miRNA data are not available in our cohorts, we foresee the importance of miRNA analysis for further understanding mechanisms of pathway dysregulation, especially when applies to therapeutic resistance.[100-102] The presence of miRNAs in tumor-derived exosomes has recently been postulated to play important roles in facilitating metastasis, and this work suggests that exosomes containing tumor-derived miRNAs which regulate one of these five pathways may also play a role in the spread of tamoxifen resistance. [103]

In addition, availability of single-cell profiles for investigation of therapeutic response has proven to be invaluable[104] in understanding of therapeutic targets for complex diseases, including cancer. Thus, as such profiles become available, we foresee their immediate utilization for elucidation of mechanisms of primary and secondary therapy resistance, and we investigate the miRNAs which are known to regulate any of the 5 molecular pathways.

Chapter V

CONCLUSIONS

In conclusion, we have demonstrated that a systematic computational pathway-centric method could identify molecular pathways to predict tamoxifen resistance, including *(i)* Retrograde Neurotrophin Signalling, *(ii)* Loss of NLP from Mitotic Centrosomes, *(iii)* RNA Polymerase III Transcription Initiation from Type 2 Promoter, *(iv)* EIF2 pathway, and *(v)* Valine Leucine and Isoleucine Biosynthesis. We propose that our finding can be ultimately utilized to prioritize and determine *(i)* cases at higher risk of developing resistance to tamoxifen that should be considered for alternative treatment manipulations (for instance, alternative endocrine therapy, radiation therapy, or chemotherapy etc.) and *(ii)* cases who would benefit maximally from tamoxifen therapy.

REFERENCES

1. Zhang MH, Man HT, Zhao XD, Dong N, Ma SL. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomed Rep.* 2014;2(1):41-52.
2. Pedraza V, Gomez-Capilla JA, Escaramis G, et al. Gene expression signatures in breast cancer distinguish phenotype characteristics, histologic subtypes, and tumor invasiveness. *Cancer.* 2010;116(2):486-496.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA: a cancer journal for clinicians.* 2019;69(1):7-34.
4. Chang M. Tamoxifen resistance in breast cancer. *Biomol Ther (Seoul).* 2012;20(3):256-267.
5. Hayes EL, Lewis-Wambi JS. Mechanisms of endocrine resistance in breast cancer: an overview of the proposed roles of noncoding RNA. *Breast Cancer Res.* 2015;17:40.
6. Group EBCTC. Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet.* 1998;351(9114):1451-1467.
7. Hackshaw A, Roughton M, Forsyth S, et al. Long-term benefits of 5 years of tamoxifen: 10-year follow-up of a large randomized trial in women at least 50 years of age with early breast cancer. *J Clin Oncol.* 2011;29(13):1657-1663.
8. Davies C, Godwin J, Gray R, et al. Early Breast Cancer Trialists' Collaborative G. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet.* 2011;378(9793):771-784.
9. Davies C, Pan H, Godwin J, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *The Lancet.* 2013;381(9869):805-816.
10. Loi S, Haibe-Kains B, Desmedt C, et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics.* 2008;9:239.
11. Gallo MA, Kaufman D. Antagonistic and agonistic effects of tamoxifen: significance in human cancer. Paper presented at: Seminars in oncology1997.
12. Fox EM, Arteaga CL, Miller TW. Abrogating endocrine resistance by targeting ERalpha and PI3K in breast cancer. *Front Oncol.* 2012;2:145.
13. Osborne CK. Tamoxifen in the treatment of breast cancer. *New England Journal of Medicine.* 1998;339(22):1609-1618.
14. Shou J, Massarweh S, Osborne CK, et al. Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. *J Natl Cancer Inst.* 2004;96(12):926-935.
15. Osborne CK, Bardou V, Hopp TA, et al. Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst.* 2003;95(5):353-361.
16. Cui X, Schiff R, Arpino G, Osborne CK, Lee AV. Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *Journal of clinical oncology.* 2005;23(30):7721-7735.
17. Dowsett M, Allred C, Knox J, et al. Relationship between quantitative estrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER-2) status with recurrence in the Arimidex, Tamoxifen, Alone or in Combination trial. *Journal of clinical oncology.* 2008;26(7):1059-1065.

18. Men X, Ma J, Wu T, et al. Transcriptome profiling identified differentially expressed genes and pathways associated with tamoxifen resistance in human breast cancer. *Oncotarget.* 2018;9(3):4074-4089.
19. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004;351(27):2817-2826.
20. Ma X-J, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer cell.* 2004;5(6):607-616.
21. Chen J, Wang Y, Shen B, Zhang D. Molecular signature of cancer at gene level or pathway level? Case studies of colorectal cancer and prostate cancer microarray data. *Computational and mathematical methods in medicine.* 2013;2013.
22. Wang Y, Chen J, Li Q, et al. Identifying novel prostate cancer associated pathways based on integrative microarray data analysis. *Computational biology and chemistry.* 2011;35(3):151-158.
23. Myers JS, von Lersner AK, Robbins CJ, Sang Q-XA. Differentially expressed genes and signature pathways of human prostate cancer. *PloS one.* 2015;10(12):e0145322.
24. Abraham G, Kowalczyk A, Loi S, Haviv I, Zobel J. Prediction of breast cancer prognosis using gene set statistics provides signature stability and biological context. *BMC bioinformatics.* 2010;11(1):277.
25. Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ. Discovering statistically significant pathways in expression profiling studies. *Proceedings of the National Academy of Sciences.* 2005;102(38):13544-13549.
26. Lee E, Chuang H-Y, Kim J-W, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. *PLoS computational biology.* 2008;4(11):e1000217.
27. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Molecular systems biology.* 2007;3(1):140.
28. Epsi NJ, Panja S, Pine SR. pathCHEMO, a generalizable computational framework uncovers molecular pathways of chemoresistance in lung adenocarcinoma. 2019;2:334.
29. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *Journal of clinical oncology.* 2007;25(10):1239.
30. Zhong Q, Fang J, Huang Z, et al. A response prediction model for taxane, cisplatin, and 5-fluorouracil chemotherapy in hypopharyngeal carcinoma. *Scientific reports.* 2018;8(1):12675.
31. Yu K, Sang Q-XA, Lung P-Y, et al. Personalized chemotherapy selection for breast cancer using gene expression profiles. *Scientific reports.* 2017;7:43294.
32. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 2013;41(Database issue):D991-995.
33. ThermoFisher Scientific. Human genome u133 set - support materials. 2018, May 25; www.affymetrix.com/support/technical/byproduct.affx?product=hgu133.
34. Negi SK, Guda C. Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Sci Rep.* 2017;7(1):897.
35. Arnatkevic Iute A, Fulcher BD, Fornito A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage.* 2019;189:353-367.
36. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009;27(8):1160-1167.
37. Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res.* 2012;18(16):4465-4472.

38. Haibe-Kains B, Schroeder M, Bontempi G, Sotiriou C, Quackenbush J. genefu: Relevant functions for gene expression analysis, especially in breast cancer. *R/Bioconductor version: Development (212).* 2011.
39. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. *J Mol Diagn.* 2003;5(2):73-81.
40. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol.* 2008;4(11):e1000217.
41. Fabregat A, Sidiropoulos K, Garapati P, et al. The reactome pathway knowledgebase. *Nucleic acids research.* 2015;44(D1):D481-D487.
42. Pandey R, Guru RK, Mount DW. Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. *Bioinformatics.* 2004;20(13):2156-2158.
43. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research.* 1999;27(1):29-34.
44. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-15550.
45. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* 2009;462(7269):108-112.
46. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological).* 1972;34(2):187-202.
47. Therneau TM, Grambsch PM. *Modeling survival data: extending the Cox model.* Springer Science & Business Media; 2013.
48. Maaten Lvd, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research.* 2008;9(Nov):2579-2605.
49. Taskesen E, Reinders MJ. 2D Representation of Transcriptomes by t-SNE Exposes Relatedness between Human Tissues. *PLoS One.* 2016;11(2):e0149853.
50. Mwangi B, Soares JC, Hasan KM. Visualization and unsupervised predictive clustering of high-dimensional multimodal neuroimaging data. *J Neurosci Methods.* 2014;236:19-25.
51. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics).* 1979;28(1):100-108.
52. stat. K-means clustering. 2019, Mar 21; https://stat.ethz.ch/R-manual/R-devel/library/stats/html/kmeans.html.
53. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine.* 2013;4(2):627.
54. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.
55. Zeileis A, Kleiber C, Jackman S. Regression models for count data in R. *Journal of statistical software.* 2008;27(8):1-25.
56. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics.* 2011;12(1):77.
57. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res.* 2010;1(4):274-278.
58. Mosteller F, Tukey JW. Data analysis, including statistics. *Handbook of social psychology.* 1968;2:80-203.
59. Welch BL. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika.* 1947;34(1/2):28-35.

60. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A.* 2001;98(19):10869-10874.
61. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics.* 2006;7:96.
62. Rouzier R, Pusztai L, Delaloge S, et al. Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. *J Clin Oncol.* 2005;23(33):8331-8339.
63. Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. *Oncologist.* 2004;9(6):606-616.
64. Høst H, Lund E. Age as a prognostic factor in breast cancer. *Cancer.* 1986;57(11):2217-2221.
65. Ahn HJ, Jung SJ, Kim TH, Oh MK, Yoon H-K. Differences in clinical outcomes between luminal A and B type breast cancers according to the St. Gallen Consensus 2013. *Journal of breast cancer.* 2015;18(2):149-159.
66. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet.* 2005;365(9460):671-679.
67. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002;415(6871):530-536.
68. Menashe I, Figueroa JD, Garcia-Closas M, et al. Large-scale pathway-based analysis of bladder cancer genome-wide association data from five studies of European background. *PloS one.* 2012;7(1):e29396.
69. Kim J, Lee Y, Cho H-J, et al. NTRK1 fusion in glioblastoma multiforme. *PLoS One.* 2014;9(3):e91940.
70. Martin-Zanca D, Hughes SH, Barbacid M. A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. *Nature.* 1986;319(6056):743.
71. Greco A, Pierotti M, Bongarzone I, Pagliardini S, Lanzi C, Della GP. TRK-T1 is a novel oncogene formed by the fusion of TPR and TRK genes in human papillary thyroid carcinomas. *Oncogene.* 1992;7(2):237-242.
72. Vaishnavi A, Capelletti M, Le AT, et al. Oncogenic and drug-sensitive NTRK1 rearrangements in lung cancer. *Nature medicine.* 2013;19(11):1469.
73. Vaishnavi A, Le AT, Doebele RC. TRKing down an old oncogene in a new era of targeted therapy. *Cancer discovery.* 2015;5(1):25-34.
74. Li J, Zhan Q. The role of centrosomal Nlp in the control of mitotic progression and tumourigenesis. *British journal of cancer.* 2011;104(10):1523.
75. Jin S, Gao H, Mazzacurati L, et al. BRCA1 interaction of centrosomal protein Nlp is required for successful mitotic progression. *Journal of Biological Chemistry.* 2009;284(34):22970-22977.
76. Zhao W, Song Y, Xu B, Zhan Q. Overexpression of centrosomal protein Nlp confers breast carcinoma resistance to paclitaxel. *Cancer biology & therapy.* 2012;13(3):156-163.
77. Strebhardt K, Ullrich A. Targeting polo-like kinase 1 for cancer therapy. *Nature reviews cancer.* 2006;6(4):321.
78. Cooper G. Eukaryotic RNA Polymerases and General Transcription Factors. _____ *The Cell: A Molecular Approach 2nd ed Sunderland (MA): Sinauer Associates.* 2000.
79. Carter R, Drouin G. Structural differentiation of the three eukaryotic RNA polymerases. *Genomics.* 2009;94(6):388-396.

80. Simms TA, Dugas SL, Gremillion JC, et al. TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae. *Eukaryotic cell.* 2008;7(12):2078-2086.
81. White RJ. RNA polymerase III transcription and cancer. *oncogene.* 2004;23(18):3208-3216.
82. Han Y, Yan C, Fishbain S, Ivanov I, He Y. Structural visualization of RNA polymerase III transcription machineries. *Cell discovery.* 2018;4(1):1-15.
83. Burwick N, Aktas BH. The eIF2-alpha kinase HRI: A potential target beyond the red blood cell. *Expert opinion on therapeutic targets.* 2017;21(12):1171-1177.
84. Donze O, Jagus R, Koromilas A, Hershey J, Sonenberg N. Abrogation of translation initiation factor eIF-2 phosphorylation causes malignant transformation of NIH 3T3 cells. *The EMBO Journal.* 1995;14(15):3828-3834.
85. Lobo MV, Martín ME, Pérez MI, et al. Levels, phosphorylation status and cellular localization of translational factor eIF2 in gastrointestinal carcinomas. *The Histochemical Journal.* 2000;32(3):139-150.
86. Wang S, Rosenwald IB, Hutzler MJ, et al. Expression of the eukaryotic translation initiation factors 4E and 2α in non-Hodgkin's lymphomas. *The American journal of pathology.* 1999;155(1):247-255.
87. Burwick N, Zhang MY, de la Puente P, et al. The eIF2-alpha kinase HRI is a novel therapeutic target in multiple myeloma. *Leukemia research.* 2017;55:23-32.
88. Hutson SM, Sweatt AJ, LaNoue KF. Branched-chain amino acid metabolism: implications for establishing safe intakes. *The Journal of nutrition.* 2005;135(6):1557S-1564S.
89. Hutson SM, Fenstermacher D, Mahar C. Role of mitochondrial transamination in branched chain amino acid metabolism. *Journal of Biological Chemistry.* 1988;263(8):3618-3625.
90. Wallin R, Hall TR, Hutson SM. Purification of branched chain aminotransferase from rat heart mitochondria. *Journal of Biological Chemistry.* 1990;265(11):6019-6024.
91. Hall T, Wallin R, Reinhart G, Hutson S. Branched chain aminotransferase isoenzymes. Purification and characterization of the rat brain isoenzyme. *Journal of Biological Chemistry.* 1993;268(5):3092-3098.
92. Mayers JR, Torrence ME, Danai LV, et al. Tissue of origin dictates branched-chain amino acid metabolism in mutant Kras-driven cancers. *Science.* 2016;353(6304):1161-1165.
93. Dey P, Baddour J, Muller F, et al. Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature.* 2017;542(7639):119.
94. Tönjes M, Barbus S, Park YJ, et al. BCAT1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type IDH1. *Nature medicine.* 2013;19(7):901.
95. Zhou W, Feng X, Ren C, et al. Over-expression of BCAT1, a c-Myc target gene, induces cell proliferation, migration and invasion in nasopharyngeal carcinoma. *Molecular cancer.* 2013;12(1):53.
96. Sharma D, Blum J, Yang X, Beaulieu N, Macleod AR, Davidson NE. Release of methyl CpG binding proteins and histone deacetylase 1 from the estrogen receptor α (ER) promoter upon reactivation in ER-negative human breast cancer cells. *Molecular endocrinology.* 2005;19(7):1740-1751.
97. Jin S, Zeng X, Fang J, et al. A network-based approach to uncover microRNA-mediated disease comorbidities and potential pathobiological implications. *npj Systems Biology and Applications.* 2019;5(1):41.
98. Braga TV, Evangelista FCG, Gomes LC, Araujo S, Carvalho MDG, Sabino AP. Evaluation of MiR-15a and MiR-16-1 as prognostic biomarkers in chronic lymphocytic leukemia. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie.* 2017;92:864-869.

99. Bandi N, Zbinden S, Gugger M, et al. miR-15a and miR-16 are implicated in cell cycle regulation in a Rb-dependent manner and are frequently deleted or down-regulated in non-small cell lung cancer. *Cancer research.* 2009;69(13):5553-5559.
100. Cava C, Colaprico A, Bertoli G, Bontempi G, Mauri G, Castiglioni I. How interacting pathways are regulated by miRNAs in breast cancer subtypes. *BMC bioinformatics.* 2016;17(12):348.
101. Miller TE, Ghoshal K, Ramaswamy B, et al. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *Journal of biological chemistry.* 2008;283(44):29897-29903.
102. Cimino D, De Pitta C, Orso F, et al. miR148b is a major coordinator of breast cancer progression in a relapse-associated microRNA signature by targeting ITGA5, ROCK1, PIK3CA, NRAS, and CSF1. *The FASEB Journal.* 2013;27(3):1223-1235.
103. Sun Z, Shi K, Yang S, et al. Effect of exosomal miRNA on cancer biology and clinical applications. *Molecular cancer.* 2018;17(1):147.
104. Gawel DR, Serra-Musach J, Lilja S, et al. A validated single-cell-based strategy to identify diagnostic and therapeutic targets in complex diseases. *Genome medicine.* 2019;11(1):47.