

Genomic Features of Rapid Versus Late Relapse in Triple Negative Breast Cancer

Yiqing Zhang

Ohio State University College of Medicine

Sarah Asad

Ohio State University College of Medicine

Zachary Weber

The Ohio State University

David Tallman

Ohio State University College of Medicine

William Nock

Ohio State University College of Medicine

Meghan Wyse

Ohio State University Comprehensive Cancer Center

Jerome F. Bey

Ohio State University College of Medicine

Kristin L. Dean

Ohio State University College of Medicine

Elizabeth J. Adams

Ohio State University College of Medicine

Sinclair Stockard

Ohio State University College of Medicine

Jasneet Singh

Ohio State University College of Medicine

Eric P. Winer

Dana-Farber Cancer Institute

Nancy U. Lin

Dana-Farber Cancer Institute

Yi-Zhou Jiang

Fudan University Shanghai Cancer Center

Ding Ma

Fudan University Shanghai Cancer Center

Peng Wang

Bio-med Big Data Center, CAS-MPG Partner Institute for Computational Biology, University of Chinese

Leming Shi

Fudan University

Wei Huang

Chinese National Human Genome Center at Shanghai (CHGC) and Shanghai Industrial Technology Institute (SITI)

Zhi-Ming Shao

Fudan University Shanghai Cancer Center

Mathew Cherian

Ohio State University College of Medicine

Maryam B. Lustberg

Ohio State University College of Medicine

Bhuvaneswari Ramaswamy

Ohio State University College of Medicine

Sagar Sardesai

Ohio State University College of Medicine

Jeffrey VanDeusen

Ohio State University College of Medicine

Nicole O. Williams

Ohio State University College of Medicine

Robert Wesolowski

Ohio State University College of Medicine

Samilia Obeng-Gyasi

Ohio State University College of Medicine

Gina Sizemore

Ohio State University College of Medicine

Steve Sizemore

Ohio State University College of Medicine

Claire Verschraegen

Ohio State University College of Medicine

Daniel G. Stover (✉ daniel.stover@osumc.edu)

Ohio State University College of Medicine

Research Article

Keywords: Breast Cancer, triple-negative breast cancer, machine learning

Posted Date: March 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-332432/v1>

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Triple-negative breast cancer (TNBC) is a heterogeneous disease and we have previously shown that rapid relapse of TNBC is associated with distinct sociodemographic features. We hypothesized that rapid versus late relapse in TNBC is also defined by distinct clinical and genomic features of primary tumors.

Methods: Using three publicly-available datasets, we identified 453 patients diagnosed with primary TNBC with adequate follow-up to be characterized as 'rapid relapse' (rrTNBC; relapse/death ≤ 2 years of diagnosis), 'late relapse' (lrTNBC; > 2 years) or 'no relapse' (nrTNBC: > 5 years no relapse/death). We explored basic clinical and primary tumor multi-omic data, including whole transcriptome (n=453), and whole genome copy number and mutation data for 171 cancer-related genes (n=317). Association of rapid relapse with clinical and genomic features were assessed using Pearson chi-squared tests, t-tests, ANOVA, and Fisher exact tests. We evaluated logistic regression models of clinical features with subtype versus two models that integrated significant genomic features.

Results: Relative to nrTNBC, both rrTNBC and lrTNBC had significantly lower immune signatures and immune signatures were highly correlated to anti-tumor CD8 T-cell, M1 macrophage, and gamma-delta T-cell CIBERSORT inferred immune subsets. Intriguingly, lrTNBCs were enriched for luminal signatures. There was no difference in tumor mutation burden or percent genome altered across groups. Logistic regression models that incorporate genomic features significantly outperformed standard clinical/subtype models in training (n=63 patients), testing (n=63) and independent validation (n=34) cohorts, although performance of all models were overall modest.

Conclusions: We identify clinical and genomic features associated with rapid relapse TNBC for further study of this aggressive TNBC subset.

Background

Triple negative breast cancer (TNBC) is an aggressive breast cancer subtype defined by lack of targetable estrogen receptor (ER), progesterone receptor (PR), and HER2.[1] TNBC accounts for 15% of breast cancer cases, yet is responsible for 35% of breast cancer related deaths.[1, 2] Relative to hormone receptor positive breast cancers, TNBCs are more likely to develop distant rather than local recurrence and TNBCs spread more frequently to visceral sites, including lung and brain.[2–4] Understanding determinants of distant relapse is imperative as the median overall survival after diagnosis of metastatic disease was historically only 13–17 months[2, 5] and remains only 25 months even among patients with PD-L1 positive TNBC receiving chemo-immunotherapy.[6]

Advances in sequencing technology have facilitated comprehensive molecular profiling of breast cancers, including subsets of TNBC.[7, 8] Two landmark analysis of primary TNBCs revealed six subtypes of TNBC with distinct expression profiles[9, 10] and an integrated copy number/transcriptome analysis

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js analyses demonstrate high frequency of

mutations in *TP53* (~ 75% of TNBCs) and *PIK3CA* ~ 25%[11–13] while TNBCs also reflect widespread copy number alterations.[11–13] The existing TNBC subsets/groupings provide a critical framework for understanding intrinsic genomic characteristics but are only associated with modest differences in patient survival. Among the approximately 30% of TNBCs who develop metastatic disease, a subset have an aggressive phenotype associated with rapid relapse, therapeutic resistance, and poor prognosis, while others have a relatively late relapse associated with more indolent or treatment responsive disease – yet we have a poor understanding of genomic features associated with distinct timing of relapse.[1, 2, 14]

To more accurately understand the differences in patient outcome in TNBC, we sought to understand distinct clinical and genomic features among primary TNBCs categorized based on outcome: rapid (rrTNBC), late (lrTNBC) and no relapse (nrTNBC). In several large TNBC cohort studies, the median time to distant metastasis was around 2 years, ranging from 19.7 to 31.2 months,[2, 14–16] thus we define rrTNBC as relapse or death within 24 months of diagnosis. We previously demonstrated in two large cohorts (Surveillance, Epidemiology, and End Results Program/SEER and National Comprehensive Cancer Network/NCCN) that disparities in sociodemographic features are strongly associated with rrTNBC, including insurance type, race, and surgical management.[17–19] These studies demonstrate the relevance of understanding factors contributing to rrTNBC yet are limited by lack of biologic understanding.

As an initial investigation of genomic features associated with rrTNBC, we aggregated data across multiple cohorts then utilized a train/test split and an independent validation cohort to model predictors of rapid versus late relapse.

Methods

Patient and Tumor Characteristics

Patient-specific data were obtained from The Cancer Genome Atlas (TCGA)[12], Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)[20, 21], our published meta-analysis (“neoadjuvant dataset” as described previously)[7], and the Fudan TNBC cohort.[22] These variables included age at diagnosis, grade, stage at diagnosis, pathologic receptor status (ER, PR, and HER2), response to neoadjuvant chemotherapy (when available), and distant metastasis-free or overall survival. TNBC was defined as being negative for ER, PR, and HER2: immunohistochemistry (IHC) 0 and FISH HER2/CEP17 ratio of less than 2.0. Neoadjuvant chemotherapy response was based on study-reported outcomes. As we previously reported, all patients in the “neoadjuvant dataset” received neoadjuvant chemotherapy but from diverse regimens: 41% of patients received anthracycline/taxane +/- alkylator, 15% anthracycline +/- alkylator, 35% taxane alone, and 9% anthracycline/platinum.

Genomic Data

For data from the METABRIC, normalized gene expression data, copy number data, and somatic mutation

Archive (IDs EGAD00010000210 and EGAD0001000021) and associated publications[21, 23]. Copy number segmented data files were processed using GISTIC2.0.[24] For data from TCGA, breast cancer gene expression data, GISTIC copy number data, and somatic mutation data were obtained from the XENABrowser (version 2015-02-24). Gene expression data from 17 published studies of breast cancer patients prior to NAC were re-processed from raw files, as previously described.[7] Genomic data from the Fudan TNBC study was downloaded from the National Omics Data Encyclopedia (accession OEP000155).[22]

Gene Expression Signatures, Expression-Based Subtypes, and Inferred Immune Subsets

Given gene expression data from multiple studies and disparate platforms, gene expression data for all TNBCs for each dataset (METABRIC $n = 287$, TCGA $n = 160$, neoadjuvant dataset $n = 446$) were extracted, quantile normalized within TNBCs from each study, and subsequently median centered. We evaluated summary expression metrics (e.g. signatures, intrinsic subtypes, CIBERSORT proportions). 125 published gene expression signatures were calculated as we have previously described.[7] We determined PAM50 intrinsic breast cancer subtype using the ‘Bioclassifier’ package from Parker et al. after balancing TNBC data with an equal number of ER-positive cases for each dataset.[25] TNBC subtype was determined using the TNBCtype tool.[9, 26] Proportion of infiltrating immune cell subsets were calculated using the CIBERSORT algorithm.[27]

Modeling and Performance

We compared the performance of three logistic regression models in predicting rapid relapse versus late relapse. The “null model” contained only clinical variables (age/stage at diagnosis and PAM50/TNBC subtype). The “null plus significant genomic features”, adds any feature significantly different between rrTNBC and lrTNBC with a nominal p -value < 0.05 . The “genomic features reduced”, is a reduced version of the second model that only includes features among the top 25 most important genomic features in at least half of the independent runs. Lasso reduction and tuning of the regularization parameter λ were performed. To evaluate model performance, we calculated the average receiver-operator characteristic (ROC) AUC of the 25 runs, and 95% confidence interval was calculated using the standard deviation of the sample of means.

Statistical Analysis

Differences in patient and tumor characteristics were evaluated using Pearson chi-squared tests. The association of gene signatures with neoadjuvant chemotherapy response was evaluated using simple linear regression and t-tests. All calculations of association were multiple-testing corrected using the Benjamini–Hochberg procedure for false discovery rate. For continuous variables, we calculated p -values comparing rapid vs. late and relapse vs. no relapse using ANOVA and logistic regression. For count variables (e.g. mutated vs. not) we used Fisher exact tests to evaluate relapse vs. not and rapid vs. late relapse. P -values for CIBERSORT and mutation signatures were evaluated using logistic regression, while

CNAs, and mutations were evaluated using Fisher exact tests. Data visualization was made using ggplot2.[28] All statistical analyses were performed in R version 3.4.1.

Results

Defining Rapid vs. Late vs. No Relapse Triple-Negative Breast Cancer

From three large cohorts with primary breast cancer genomic data – TCGA,[12] METABRIC,[20, 21] and our prior breast cancer gene expression meta-analysis[7] – we identified 893 TNBCs from a total of 4473 breast cancer cases. For our analyses, we included patients with at least 60 months of follow-up or those with a distant metastasis-free survival (DMFS) event prior to our 60-month cutoff, leaving a total of 453 TNBCs in our evaluable dataset. Of these, 453 had gene expression data, 317 had copy number data, and 317 had mutation data. (Fig. 1A).

We assessed the percentage of total DMFS events each year (Fig. 1B). In this dataset, over 20% of DMFS events occurred each of the first two years after diagnosis, categorized as ‘rapid relapse’ (rrTNBC). Among lrTNBCs, most DMFS events occurred within the first five years after diagnosis, with sporadic events beyond year 6. Our main goal was to identify differences among TNBCs with clinically distinct outcomes, so we visualized DMFS for our relapse categorization (Fig. 1C) in comparison with DMFS for existing intrinsic expression-based subtype approaches PAM50[25] (Fig. 1D) or Lehmann/Pietenpol TNBCtype[9] (Fig. 1E) within the same cohort. The Lehmann/Pietenpol TNBCtype (log-rank $p = 0.01$), but not PAM50, was associated with significant differences in DMFS. The strikingly different visualized outcomes suggests that our relapse categorization does, in fact, identify truly distinct subsets based on outcome when compared to approaches that focus on intrinsic features.

Patient and Tumor Characteristics

We evaluated the association of clinical, pathologic, and intrinsic expression subtype with rapid vs. late vs. no relapse status (Table 1). There was no significant difference in age at diagnosis or grade, however, rrTNBCs were significantly more likely to be higher stage (Chi-square $p = 1.9e-10$). The majority of patients were basal-like PAM50 subtype (78%), but, lrTNBCs were significantly more likely to be non-basal (non-basal: rrTNBC 18%, lrTNBC 29%, nrTNBC 20%, Chi-square $p = 0.03$). Lehmann/Pietenpol TNBC subtype also reflected significant differences across groups (Chi-square $p = 0.02$). The immunomodulatory phenotype was highest in nrTNBC (16% rrTNBC, 16% lrTNBC, 24% nrTNBC), luminal androgen receptor was highest in lrTNBC (9% rrTNBC, 16% lrTNBC, 9% nrTNBC), and basal-like 2 was highest in rrTNBC (15% rrTNBC, 9% lrTNBC, 6% nrTNBC). A subset of patients in this cohort (127/453; 28.0%) had data on response to neoadjuvant chemotherapy (NAC). As anticipated, those patients with rrTNBC or lrTNBC were significantly more likely to have residual disease (RD) after neoadjuvant chemotherapy (93% and 94% RD, respectively), relative to those with nrTNBC (51% RD; Chi-square $p =$

1.9e-7). Intriguingly, the rate of residual disease was similar among rrTNBC and lrTNBC despite markedly different timing of relapse.

Table 1
Cohort clinical and pathologic features

Rapid Relapse	All patients n = 453	Rapid Relapse n = 110	Late Relapse n = 125	No Relapse n = 218	P
	n (%)	n (%)	n (%)	n (%)	
Age at diagnosis, by decade					0.12
< 40 years	75 (17)	15 (14)	27 (22)	33 (15)	
40 to 50 years	117 (26)	27 (24)	28 (22)	62 (29)	
50 to 60 years	124 (27)	33 (30)	25 (20)	66 (30)	
> 60 years	137 (30)	35 (32)	45 (36)	57 (26)	
Grade at diagnosis					0.86
I	6 (2)	2 (2)	2 (2)	2 (1)	
II	54 (14)	14 (15)	17 (26)	23 (13)	
III	321 (84)	75 (83)	89 (82)	157 (86)	
Stage at diagnosis					< 0.001
I	73 (17)	3 (3)	18 (15)	52 (25)	
II	231 (54)	43 (44)	69 (56)	119 (58)	
III	123 (29)	52 (53)	35 (29)	36 (27)	
Pam50 Subtype					0.03
Basal	354 (78)	90 (82)	89 (71)	175 (80)	
Non-Basal	99 (22)	20 (18)	36 (29)	43 (20)	
TNBC Subtype					0.02
Basal-like 1	103 (23)	23 (21)	26 (21)	54 (25)	
Basal-like 2	40 (9)	16 (15)	11 (9)	13 (6)	
Immunomodulatory	91 (20)	18 (16)	20 (16)	53 (24)	
Luminal androgen receptor	49 (11)	10 (9)	20 (16)	19 (9)	
Mesenchymal	66 (14)	19 (17)	20 (16)	27 (12)	
Mesenchymal stem-like	24 (5)	10 (9)	3 (2)	11 (5)	

Rapid Relapse	All patients n = 453	Rapid Relapse n = 110	Late Relapse n = 125	No Relapse n = 218	P
	n (%)	n (%)	n (%)	n (%)	
Unselected	80 (18)	14 (13)	25 (20)	41 (19)	

Response to Neoadjuvant Chemotherapy and Survival in TNBC: Immune and Expression Signatures

Response to NAC is known to be a robust prognostic biomarker in TNBC.[29] In this cohort, only 28% (127/453) of patients received NAC and many of the regimens were non-standard (e.g. taxane alone). Because of this, the pathologic complete response (pCR) after NAC was only 22.8%, much lower than modern current regimens, typically ~ 40%. Despite these significant limitations, pCR was strongly associated with nrTNBC ($p < 0.001$). The patients with data on response to NAC all had whole transcriptome data but no available mutation or copy number data, so we calculated a score for 125 published gene expression signatures and evaluated the association of each signature with NAC response (pCR vs. RD) and DMFS. Signatures were grouped by phenotype as previously described[7] ($n = 127$ patients; Fig. 2A). Immune signatures were associated with better prognosis and most were also associated with improved response to NAC. Proliferation signatures tended to be associated with improved response to NAC, as we have previously described[7], yet there was variable association with DMFS.

To understand what immune cell types in the tumor microenvironment may be reflected by the immune signatures, we visualized the association of three representative signatures from each group (immune, proliferation, ER/HER2, mesenchymal) with the relative proportion of 22 inferred immune cell subsets via CIBERSORT (Fig. 2B).[27] Immune signatures were strongly positively correlated with anti-tumor immune cell types including M1 macrophages, CD8 T-cells, and memory B-cells (all Pearson's $r \geq 0.3$, all $p < 1.2e-8$) and anti-correlated with immune suppressive cell types including M2 macrophages, memory resting CD4 T-cells, resting NK cells, and resting mast cells. ER/HER2 signatures reflected an almost opposite pattern to immune signatures, with positive correlation to immune suppressive cell types and anti-correlation with anti-tumor immune cell type. Metabolic signatures appeared to have a strong correlation specifically with M0 macrophages (all Pearson's $r > 0.27$, all $p < 8.4e-9$). As a sensitivity analysis, we evaluated the association of three representative signatures from each group with 7 immune cell-type specific signatures from MSigDB[30, 31] (instead of CIBERSORT) and found similar results (Supplementary Fig. 1A).

Expression Signatures in Rapid vs. Late vs. No Relapse TNBC

To assess pathways and phenotypes associated with rapid vs. late vs. no relapse, a score was calculated for 125 published gene expression signatures across the entire dataset (Supplementary Fig. 1B).

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js groups revealed 16 signatures that were

significantly different (ANOVA FDR $p < 0.05$; Fig. 3, **Supplementary Fig. 2A-B**). Among these, five signatures were immune-related[9, 32–34] and all were significantly higher in nrTNBC relative to rrTNBC and lrTNBC. Eight significant signatures were related to luminal phenotype – all were highest in lrTNBC, lowest in rrTNBC, and intermediate in nrTNBC. Most CIBERSORT immune subsets were not statistically significant (**Supplementary Fig. 2C**), however, neutrophils were significantly higher in rrTNBC (ANOVA FDR $p = 0.001$).

Mutations and Copy Number Alterations

In this cohort, 70% (317/453) of patients had data on single nucleotide variant/mutation data including 171 cancer-related genes and whole genome CNAs.[21] Only a small subset of patients (11.7%; 53/453) had whole exome mutation data, so we focused on the 171 cancer-related genes to ensure adequate statistical power. When evaluating general mutational features, there was no significant difference in mutations per megabase (ANOVA $p = 0.64$; Fig. 4A) nor percent genome altered by copy number (ANOVA $p = 0.96$; Fig. 4B).

We first compared the frequency of alteration for each mutation and cytoband (for CNAs) for relapse (rrTNBC + lrTNBC) vs. nrTNBC (Fig. 4C) because of low mutation frequency for most genes. There were no genes that were significantly different after multiple testing (**Supplementary Fig. 3A**) when comparing relapse vs. no relapse, but *PIK3CA* mutations were more frequent in relapse relative to nrTNBC. In addition, *PTEN*, *ARID1A*, and *RYR2* mutations were enriched in nrTNBC relative to rrTNBC (Fisher exact nominal $p < 0.05$). We then compared rrTNBC vs. lrTNBC (Fig. 4D) and found that rrTNBC were significantly more likely to harbor a mutation in *TP53* compared to lrTNBC patients (Fisher exact FDR $p = 0.009$). Among CNAs, the copy number landscape was similar across the rapid vs. late vs. no relapse groups (**Supplementary Fig. 3B**) and there were no significantly altered genes or regions among these three groups after multiple test correction yet there were several regions that demonstrated enrichment within specific groups (nominal $p < 0.05$; Fig. 4C-D).

Clinical and multi-omic model of rapid vs. late relapse in TNBC

Having identified discrete clinical, expression, immune, mutation, and copy number features among primary TNBCs with distinct clinical outcomes, we sought to develop an optimal, multi-omic predictive model for rrTNBC vs. lrTNBC. We compared performance of three logistic regression models with lasso reduction (detailed in the Methods; Fig. 5A). The clinical, “null model”, performed marginally in both the testing cohort and the independent validation cohort (average AUC 0.574 and 0.525, respectively). The other two models (clinical + genomic and reduced genomic) had significantly improved performance in both the testing cohort and the independent validation (average AUC: 0.774 and 0.821 for testing; 0.645 and 0.620 for validation; Fig. 5B; all Wilcoxon rank sum $p < 0.005$). The genomic features that contributed most included clinical features (stage, expression subtypes), mutations (*ARID2*, *DNAH11*, *SETDB1*), copy number alterations (loss *LAMA2*, *CLK3*, *MLLT4*, *SYNE1* and gain *DNAH5*, *LIFR*, *PETN*), and expression

signatures (signatures of RBBP8[35], ER negative chemoresistance[36], PTEN deletion[37], beta catenin[38], STAT3[39], and RAS pathway activation[40]). We evaluated additional models, including machine learning approaches (random forest, support vector machine) as well as the universe of available genomic data, however, these additional modeling approaches were characterized by overfitting even in the context of model tuning and demonstrated no significant improvement in performance relative (data not shown).

Discussion

We previously demonstrated in two large cohorts that disparities in sociodemographic features are strongly associated with rrTNBC,[17–19] and in this report sought to investigate genomic features associated with rrTNBC. We aggregated data from the available cohorts that have multi-omic data as well as adequate follow-up to characterize TNBCs as rrTNBC, lrTNBC, or nrTNBC. Although limited by the retrospective nature and limited number of patients who received modern NAC, we provide initial observations regarding genomic features and rrTNBC.

We sought to identify distinguishing features and determine if predictive models incorporating clinical, expression-based subtype, and/or multi-omic models could identify patients at high risk of rapid relapse. lrTNBCs are more likely to be non-basal (primarily luminal A/B) and our data identify eight luminal signatures are associated with late relapse. Multiple groups have identified a ‘luminal androgen receptor’ subset of TNBC based on molecular classifications,[9, 11] and 40% (20/49) of the Lehmann LAR subtype tumors in our cohort ultimately had late relapse. To develop predictive models, we first identified the relatively few specific features that were significantly different across subsets (61 features from > 35,000 initial data points) then built models based on *a priori* feature identification. This approach led to overall good performance of multiple models, and importantly allows us to understand what genomic features contribute most. More complex modeling approaches (e.g. machine learning algorithms) did not improve model performance and led to challenges with overfitting. Our models were evaluated in over two times the number of TNBC patients available in TCGA[12] alone – a remarkable number for a disease that accounts for only approximately 15% of breast cancers.[1, 2] Collectively, our data support the categorization by Burstein et al.[11] and suggest that lrTNBCs are enriched for luminal phenotypes while rrTNBCs are likely enriched for the ‘basal-like immune suppressed’ phenotype.

Stage at diagnosis was strongly associated with rrTNBC in univariate analyses and in logistic regression models. One hypothesis is that stage at diagnosis captures non-biological features including socioeconomic or demographics features.[41–43] Race/ethnicity is complex,[44, 45] was largely unavailable in the included datasets, and warrants further study.[46, 47] In a parallel study, we investigated the association of sociodemographic features with rrTNBC among 3016 primary TNBCs at ten academic cancer centers.[48] In this large cohort, we found that stage at diagnosis remained significant, as well as Medicaid/indigent insurance, lower income, and younger age.[48] Collectively, this suggests that timing of relapse is impacted by a complex set of clinical, genomic, and sociodemographic

Response to neoadjuvant chemotherapy remains the best prognostic biomarker for TNBC,[29] but there are clear differences in disease course among TNBCs who develop relapse earlier vs. later. At the time of these analyses, no large multi-omic dataset including NAC and long-term outcomes were available although this is anticipated in the future. Despite significant limitations of NAC analyses, somewhat unexpectedly patients destined for rrTNBC and lrTNBC in this cohort had similarly high rates of residual disease to neoadjuvant chemotherapy. Both rrTNBC and lrTNBC had lower expression of immune signatures compared with nrTNBCs, reflecting reduced anti-tumor immune response. This supports our and others' work,[7, 49–53] including our analyses of the BrighTNess phase III clinical trial, which provides largest transcriptome dataset and association with NAC and demonstrated that stratifying patients by proliferation and immune signatures can effectively stratify likelihood of pCR irrespective of NAC regimen. Given the recent FDA approval of immunotherapy for metastatic TNBC[6], there is great interest to augment the existing host anti-tumor immune response.[54–57]

While this study presents promising methods to categorize TNBC relapse it does possess significant limitations. The lack of available robust multi-omic datasets with long-term outcome data leads to inherent limitations of aggregating multiple datasets. We incorporated genomic data from multiple studies, generated using multiple platforms, and over multiple years. While we have attempted to account for this through standard normalization approaches and analysis only of summary statistics (e.g. expression signatures not individual genes), batch/platform effects and computational analyses could impact our results. In addition, therapy for TNBC has changed, including incorporation of capecitabine for RD based on CREATE-X[58] and recent FDA approval of immunotherapy for metastatic, PD-L1 positive TNBC.[6]

In conclusion, we provide evidence that rrTNBC reflects a distinct clinical entity characterized by unique genomic features. Predictive modeling using clinical and genomic features in these datasets revealed modest results, but with improved data may identify patients at high risk for 'rapid relapse.' Multi-level analyses of the interaction between clinical, multi-omic, and sociodemographic features and timing of relapse are warranted.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of The Ohio State University.

Consent for publication

Not Applicable.

Availability of data and materials

Data that support the findings of this study have been deposited in the following repositories.

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Neoadjuvant dataset: Raw gene expression data and paired clinical feature data were obtained from NCBI Gene Expression Omnibus (GEO) via accession numbers GSE8465, GSE16446, GSE18728, GSE19697, GSE20194, GSE20271, GSE21974, GSE21997, GSE22093, GSE22226, GSE22358, GSE22513, GSE23988, GSE25066, GSE28796, and GSE32646).

METABRIC: Molecular Taxonomy of Breast Cancer International Consortium. Normalized gene expression data, copy number data, and paired clinical feature data were obtained from the publicly available European Genome-phenome Archive (IDs EGAD00010000210 and EGAD0001000021) with mutation data from Pereira, et al. 2016

TCGA: Normalized gene expression data, copy number data, single nucleotide variant, and paired clinical feature data TCGA data (2015-02-24 datafreeze) were obtained from the University of California, Santa Cruz cancer browser (<https://xenabrowser.net/heatmap/>).

Fudan dataset: Data can be viewed in The National Omics Data Encyclopedia (NODE) (<http://www.biosino.org/node>), accession OEP000155. Microarray data and sequence data are available in NCBI GO (OncoScan array; GEO: GSE118527) and Sequence Read Archive (WES and RNA-seq; SRA: SRP157974).

Competing Interests

E.P.W. has received research grants from Genentech and Roche. N.U.L. has received research grants from Genentech, Array Biopharma, GlaxoSmithKline, Kadmon and Novartis. R.W. has received research support from Acerta and Astra Zeneca and served on advisory boards for PUMA and Pfizer.

Funding

This research was supported by Susan G. Komen Career Catalyst Research Award (D.G.S.), Pelotonia Young Investigator Award (D.G.S.), and the Stephanie Spielman Fund (D.G.S.).

Authors' Contributions

Y.Z., W.N., E.P.W., N.U.L, and D.G.S. envisioned the project and designed the work. Y.Z.J., D.M., P.W., L.S., W.H., Z.M.S. generated genomic data for the independent validation dataset. Y.Z., W.N., M.W., Z.W, E.J.A., S.A., D.G.S. wrote the manuscript with input from S.S., D.T., J.T., M.C., M.B.L., B.R., S.S., J.V.D., N.W., AND R.W. All authors have read, revised, and approved the final manuscript. As corresponding author, D.G.S ensures that data, materials, and code comply with transparency and reproducibility standards of the field and journal; ensures that original data/materials/code upon which the submission is based are preserved following best practices in the field so that they are retrievable for reanalysis; confirms that data/materials/code presentation accurately reflects the original; and minimizes obstacles to the sharing of data/materials/code described in the work.

Acknowledgements

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

The authors thank Catherine Carson CNP, Celia Garr RN, Ashley Little, and Katherine Weber RN for clinical support making this research possible.

References

1. 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* 2007, **109**(9):1721-1728.
2. Lin NU, Claus E, Sohl J, Razzak AR, Arnaout A, Winer EP: **Sites of distant recurrence and clinical outcomes in patients with metastatic triple-negative breast cancer: high incidence of central nervous system metastases.** *Cancer* 2008, **113**(10):2638-2645.
3. Lin NU, Vanderplas A, Hughes ME, Theriault RL, Edge SB, Wong YN, Blayney DW, Niland JC, Winer EP, Weeks JC: **Clinicopathologic features, patterns of recurrence, and survival among women with triple-negative breast cancer in the National Comprehensive Cancer Network.** *Cancer* 2012, **118**(22):5463-5472.
4. Haffty BG, Yang Q, Reiss M, Kearney T, Higgins SA, Weidhaas J, Harris L, Hait W, Toppmeyer D: **Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer.** *J Clin Oncol* 2006, **24**(36):5652-5657.
5. Kassam F, Enright K, Dent R, Dranitsaris G, Myers J, Flynn C, Fralick M, Kumar R, Clemons M: **Survival outcomes for patients with metastatic triple-negative breast cancer: implications for clinical practice and trial design.** *Clinical breast cancer* 2009, **9**(1):29-33.
6. Schmid P, Adams S, Rugo HS, Schneeweiss A, Barrios CH, Iwata H, Dieras V, Hegg R, Im SA, Shaw Wright G *et al*: **Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer.** *N Engl J Med* 2018, **379**(22):2108-2121.
7. Stover DG, Coloff JL, Barry WT, Brugge JS, Winer EP, Selfors LM: **The Role of Proliferation in Determining Response to Neoadjuvant Chemotherapy in Breast Cancer: A Gene Expression-Based Meta-Analysis.** *Clin Cancer Res* 2016, **22**(24):6039-6050.
8. Balko JM, Giltmane JM, Wang K, Schwarz LJ, Young CD, Cook RS, Owens P, Sanders ME, Kuba MG, Sanchez V *et al*: **Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets.** *Cancer discovery* 2014, **4**(2):232-245.
9. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA: **Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies.** *J Clin Invest* 2011, **121**(7):2750-2767.
10. Lehmann BD, Jovanovic B, Chen X, Estrada MV, Johnson KN, Shyr Y, Moses HL, Sanders ME, Pietenpol JA: **Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection.** *PLoS One* 2016, **11**(6):e0157368.

11. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, Savage MI, Osborne CK, Hilsenbeck SG, Chang JC *et al*: **Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer.** *Clin Cancer Res* 2015, **21**(7):1688-1698.
12. The Cancer Genome Atlas: **Comprehensive molecular portraits of human breast tumours.** *Nature* 2012, **490**(7418):61-70.
13. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**(7403):346-352.
14. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA: **Triple-negative breast cancer: clinical features and patterns of recurrence.** *Clin Cancer Res* 2007, **13**(15 Pt 1):4429-4434.
15. van Roozendaal LM, Smit LHM, Duijsens G, de Vries B, Siesling S, Lobbes MBI, de Boer M, de Wilt JHW, Smidt ML: **Risk of regional recurrence in triple-negative breast cancer patients: a Dutch cohort study.** *Breast Cancer Res Treat* 2016, **156**(3):465-472.
16. Ghosh J, Agarwal S, Ganguly S, Dabkara D, Biswas B, Nandi S, Chakraborty S, Datta S, Ahmed R, Chatterjee S: **Patterns of recurrence in triple negative breast cancer patients (automated IHC) : An Indian Tertiary Care Center data.** *Journal of Clinical Oncology* 2018, **36**(15_suppl):e13128-e13128.
17. Obeng-Gyasi S, Asad S, Fisher JL, Rahrkar S, Stover DG: **Socioeconomic and Surgical Disparities are Associated with Rapid Relapse in Patients with Triple-Negative Breast Cancer.** *Ann Surg Oncol* 2021.
18. Obeng-Gyasi S, Stover DG: **ASO Author Reflections: Disparities in Oncologic Surgery: The Problem We All Live with.** *Ann Surg Oncol* 2021.
19. Asad S, Barcenas CH, Bleicher RJ, Cohen AL, Javid SH, Levine EG, Lin NU, Moy B, Niland J, Wolff AC *et al*: **Sociodemographic Factors Associated With Rapid Relapse in Triple-Negative Breast Cancer: A Multi-Institution Study.** *J Natl Compr Canc Netw* 2021:1-8.
20. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G *et al*: **The clonal and mutational evolution spectrum of primary triple-negative breast cancers.** *Nature* 2012, **486**(7403):395-399.
21. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ *et al*: **The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes.** *Nature communications* 2016, **7**:11479.
22. Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, Xiao Y, Yu KD, Liu YR, Yu Y *et al*: **Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies.** *Cancer Cell* 2019, **35**(3):428-440 e425.
23. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**(7403):346-352.

24. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G: **GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.** *Genome Biol* 2011, **12**(4):R41.
25. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**(8):1160-1167.
26. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J *et al*: **Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer.** *British journal of cancer* 2012, **107**(10):1776-1782.
27. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.** *Nature methods* 2015, **12**(5):453-457.
28. Wickam H: **ggplot2: Elegant Graphics for Data Analysis.** New York, NY: Springer-Verlag; 2009.
29. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P *et al*: **Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis.** *Lancet* 2014, **384**(9938):164-172.
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *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.
31. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, Fong S, van Lookeren Campagne M, Godowski P, Williams PM *et al*: **Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data.** *Genes and immunity* 2005, **6**(4):319-331.
32. Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, Pfitzner BM, Salat C, Loi S, Schmitt WD *et al*: **Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers.** *J Clin Oncol* 2015, **33**(9):983-991.
33. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C: **Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes.** *Clin Cancer Res* 2008, **14**(16):5158-5165.
34. Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E, Solbach C, Hanker L, Ahr A, Metzler D *et al*: **T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers.** *Breast cancer research : BCR* 2009, **11**(2):R15.
35. Furuta S, Wang JM, Wei S, Jeng YM, Jiang X, Gu B, Chen PL, Lee EY, Lee WH: **Removal of BRCA1/CtIP/ZBRK1 repressor complex on ANG1 promoter leads to accelerated mammary tumor growth contributed by prominent vasculature.** *Cancer Cell* 2006, **10**(1):13-24.
36. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, Vidaurre T, Holmes F, Souchon E, Wang H *et al*: **A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer.** *JAMA* 2011, **305**(18):1873-1881.

37. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, Koujak S, Ferrando AA, Malmstrom P, Memeo L *et al*: **Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(18):7564-7569.
38. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A *et al*: **Oncogenic pathway signatures in human cancers as a guide to targeted therapies.** *Nature* 2006, **439**(7074):353-357.
39. Alvarez JV, Febbo PG, Ramaswamy S, Loda M, Richardson A, Frank DA: **Identification of a genetic signature of activated signal transducer and activator of transcription 3 in human tumors.** *Cancer Res* 2005, **65**(12):5054-5062.
40. Loboda A, Nebozhyn M, Klinghoffer R, Frazier J, Chastain M, Arthur W, Roberts B, Zhang T, Chenard M, Haines B *et al*: **A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors.** *BMC Med Genomics* 2010, **3**:26.
41. Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman MC: **Trends in area-socioeconomic and race-ethnic disparities in breast cancer incidence, stage at diagnosis, screening, mortality, and survival among women ages 50 years and over (1987-2005).** *Cancer Epidemiol Biomarkers Prev* 2009, **18**(1):121-131.
42. Booth CM, Li G, Zhang-Salomons J, Mackillop WJ: **The impact of socioeconomic status on stage of cancer at diagnosis and survival: a population-based study in Ontario, Canada.** *Cancer* 2010, **116**(17):4160-4167.
43. Sprague BL, Trentham-Dietz A, Gangnon RE, Ramchandani R, Hampton JM, Robert SA, Remington PL, Newcomb PA: **Socioeconomic status and survival after an invasive breast cancer diagnosis.** *Cancer* 2011, **117**(7):1542-1551.
44. Schwartz KL, Crossley-May H, Vigneau FD, Brown K, Banerjee M: **Race, socioeconomic status and stage at diagnosis for five common malignancies.** *Cancer Causes Control* 2003, **14**(8):761-766.
45. Lantz PM, Mujahid M, Schwartz K, Janz NK, Fagerlin A, Salem B, Liu L, Deapen D, Katz SJ: **The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis.** *Am J Public Health* 2006, **96**(12):2173-2178.
46. Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, Yoshimatsu TF, Pitt JJ, Hoadley KA, Troester M *et al*: **Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas.** *JAMA Oncol* 2017, **3**(12):1654-1662.
47. Pitt JJ, Riester M, Zheng Y, Yoshimatsu TF, Sanni A, Oluwasola O, Veloso A, Labrot E, Wang S, Odetunde A *et al*: **Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features.** *Nature communications* 2018, **9**(1):4181.
48. Asad S, Barcenas CH, Bleicher RJ, Cohen AL, Javid SH, Levine EG, Lin NU, Moy B, Niland J, Wolff AC *et al*: **Abstract P5-07-02: Factors associated with rapid relapse in triple negative breast cancer: A**

- multi-institution study.** *Cancer Research* 2020, **80**(4 Supplement):P5-07-02-P05-07-02.
49. Salgado R, Denkert C, Campbell C, Savas P, Nuciforo P, Aura C, de Azambuja E, Eidtmann H, Ellis CE, Baselga J *et al*: **Tumor-Infiltrating Lymphocytes and Associations With Pathological Complete Response and Event-Free Survival in HER2-Positive Early-Stage Breast Cancer Treated With Lapatinib and Trastuzumab: A Secondary Analysis of the NeoALTTO Trial.** *JAMA Oncol* 2015, **1**(4):448-454.
50. Mao Y, Qu Q, Zhang Y, Liu J, Chen X, Shen K: **The value of tumor infiltrating lymphocytes (TILs) for predicting response to neoadjuvant chemotherapy in breast cancer: a systematic review and meta-analysis.** *PLoS One* 2014, **9**(12):e115103.
51. Salgado R, Denkert C, Campbell C, Savas P, Nuciforo P, Aura C, de Azambuja E, Eidtmann H, Ellis CE, Baselga J *et al*: **Tumor-Infiltrating Lymphocytes and Associations With Pathological Complete Response and Event-Free Survival in HER2-Positive Early-Stage Breast Cancer Treated With Lapatinib and Trastuzumab: A Secondary Analysis of the NeoALTTO Trial.** *JAMA Oncol* 2015, **1**(4):448-454.
52. Denkert C, Loibl S, Noske A, Roller M, Muller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C *et al*: **Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer.** *J Clin Oncol* 2010, **28**(1):105-113.
53. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E *et al*: **Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98.** *J Clin Oncol* 2013, **31**(7):860-867.
54. Telli ML, Stover DG, Loi S, Aparicio S, Carey LA, Domchek SM, Newman L, Sledge GW, Winer EP: **Homologous recombination deficiency and host anti-tumor immunity in triple-negative breast cancer.** *Breast Cancer Res Treat* 2018, **171**(1):21-31.
55. Adams S, Gatti-Mays ME, Kalinsky K, Korde LA, Sharon E, Amiri-Kordestani L, Bear H, McArthur HL, Frank E, Perlmutter J *et al*: **Current Landscape of Immunotherapy in Breast Cancer: A Review.** *JAMA Oncol* 2019.
56. Esteva FJ, Hubbard-Lucey VM, Tang J, Pusztai L: **Immunotherapy and targeted therapy combinations in metastatic breast cancer.** *The Lancet Oncology* 2019, **20**(3):e175-e186.
57. Pusztai L, Karn T, Safonov A, Abu-Khalaf MM, Bianchini G: **New strategies in breast cancer: immunotherapy.** *Clinical Cancer Research* 2016, **22**(9):2105-2110.
58. Masuda N, Lee SJ, Ohtani S, Im YH, Lee ES, Yokota I, Kuroi K, Im SA, Park BW, Kim SB *et al*: **Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy.** *N Engl J Med* 2017, **376**(22):2147-2159.
59. Cardoso F, Costa A, Senkus E, Aapro M, Andre F, Barrios CH, Bergh J, Bhattacharyya G, Biganzoli L, Cardoso MJ *et al*: **3rd ESO-ESMO international consensus guidelines for Advanced Breast Cancer (ABC 3).** *Breast* 2017, **31**:244-259.
60. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, Griffith M: **GenVisR: Genomic Visualizations in R.** *Bioinformatics* 2016, **32**(19):3012-3014.

Figures

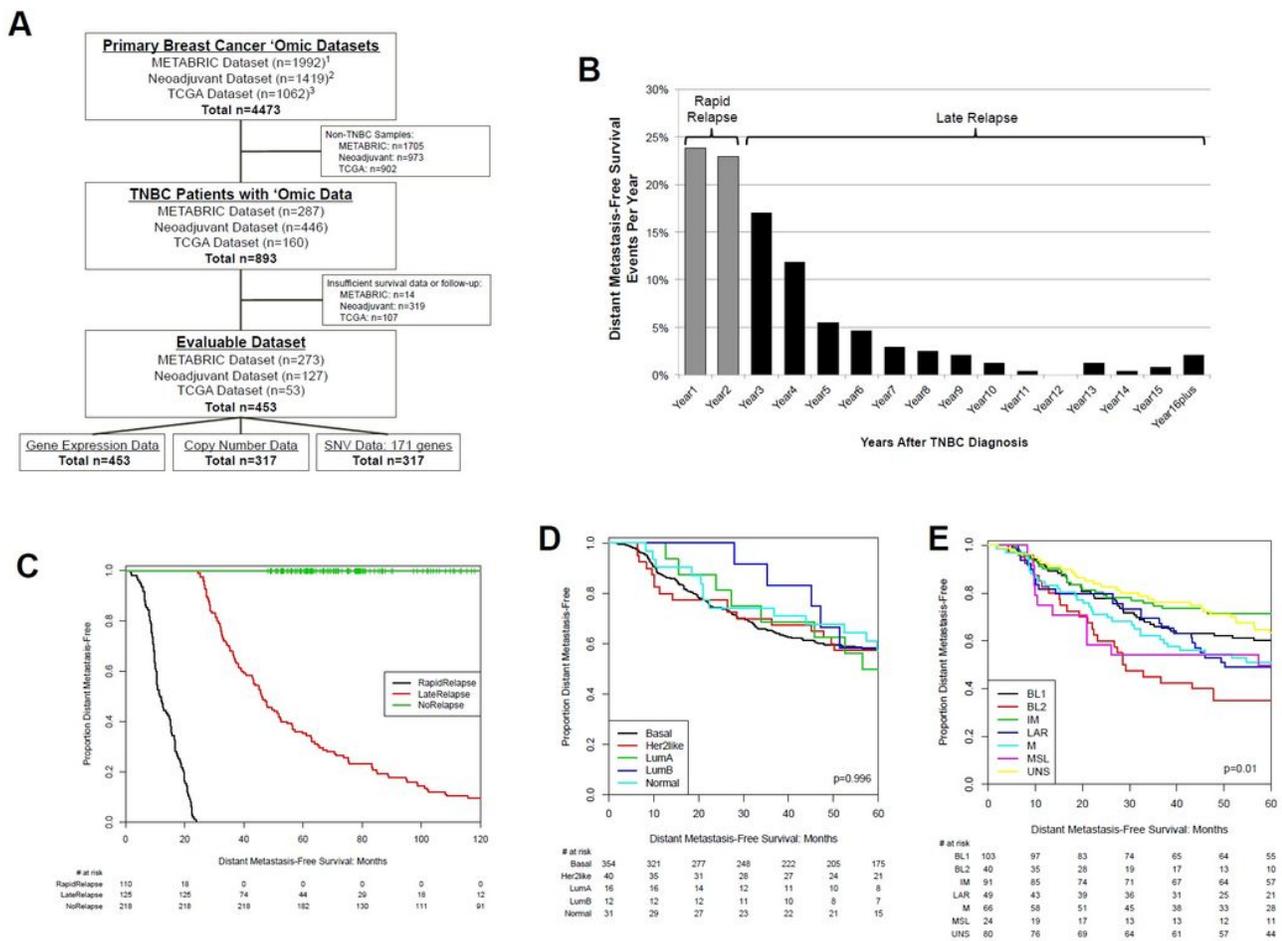


Figure 1

Study design and definition of triple-negative breast cancer (TNBC) rapid vs. late relapse. (A) REMARK diagram. (B). Proportion of distant metastasis-free survival (DMFS) events per year after diagnosis among evaluable dataset. 'Rapid relapse' was defined as DMFS events within the 2 years of diagnosis and 'late relapse' DMFS events beyond 2 years. (C-E) Kaplan-Meier diagram of DMFS in study cohort reflecting TNBC group definitions (C), compared with DMFS by intrinsic subtype approaches PAM50 subtype (D), and Lehmann TNBC subtype (E). P-value indicates log-rank test.

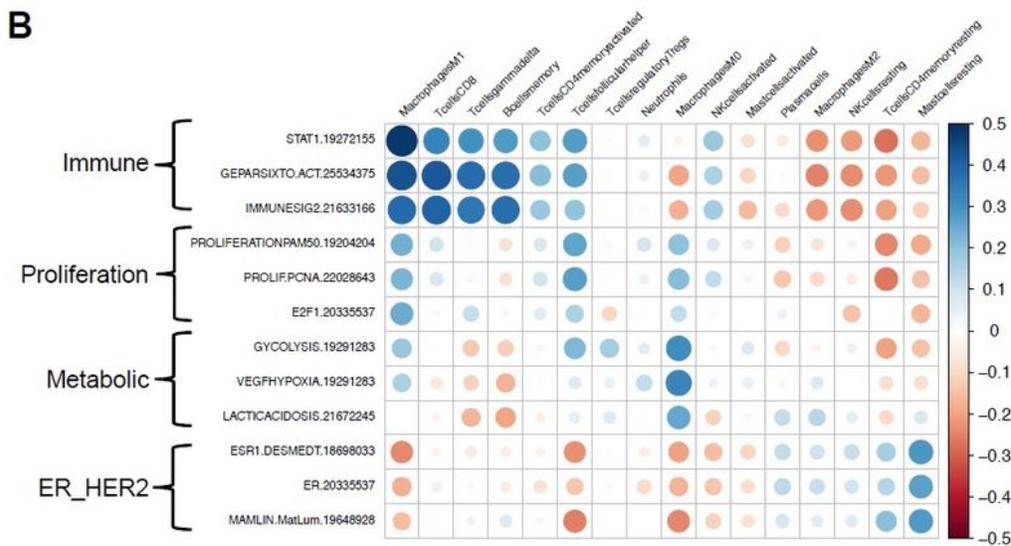
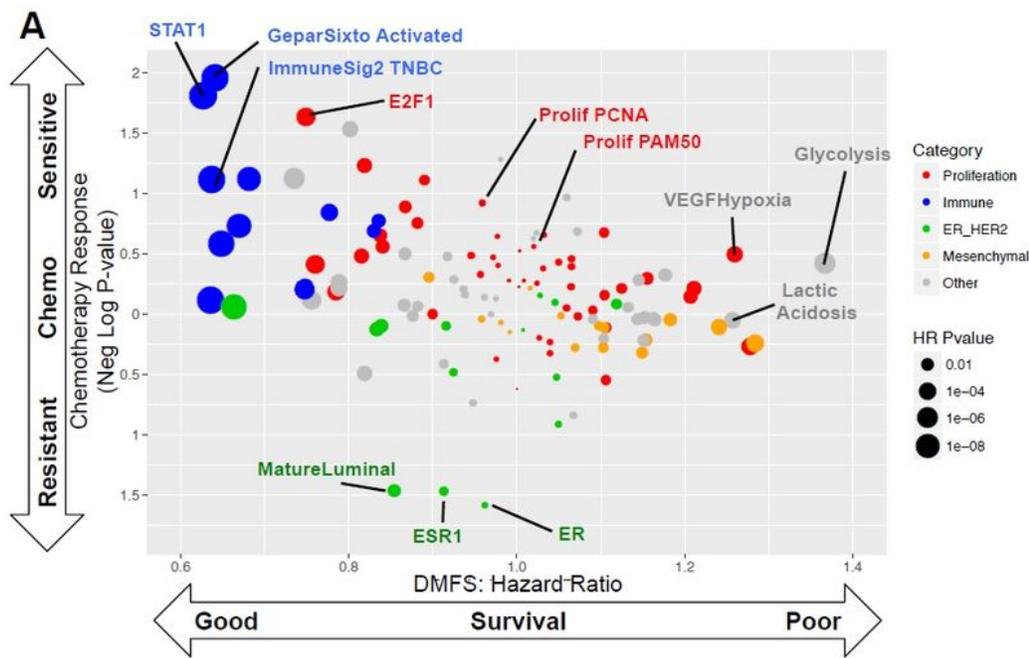
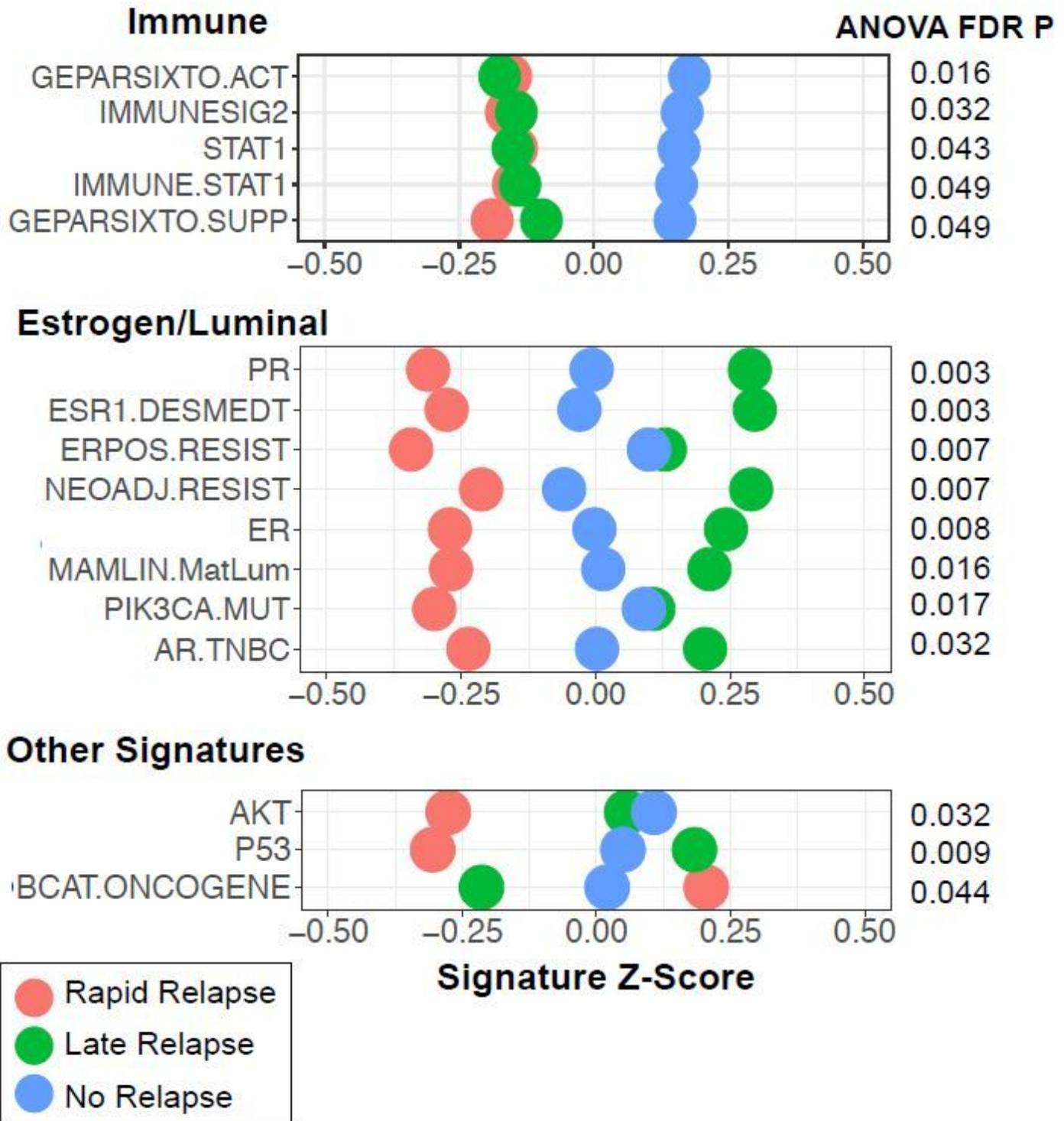


Figure 2

Immune and Expression Signatures and Response to Neoadjuvant Chemotherapy and Survival in TNBC. (A) The calculated score for 125 published gene expression signatures for 127 patients with data on response to neoadjuvant chemotherapy and distant metastasis-free survival (DMFS). Each signature is a point. The association of each signature with neoadjuvant chemotherapy response (pathologic complete

hazard ratio for each signature using DMFS (x-

axis) are displayed. Signatures were grouped by phenotype (as previously described[7]), identified by color: proliferation signatures (red), immune signatures (blue), ER/HER2 signatures (green), mesenchymal signatures (orange), others (grey). Size of each point relates to the hazard ratio p-value for each signature. (B) The association of three representative signatures from each group (immune, proliferation, ER/HER2, mesenchymal) with the relative proportion of 22 inferred immune cell subsets via CIBERSORT across all samples with gene expression data (n=453) are visualized using CorrPlot.[27, 59]



Expression Signatures in Rapid vs. Late vs. No Relapse TNBC. The calculated score for 16 published gene expression signatures that demonstrated statistical significance (ANOVA FDR $p < 0.05$) comparing rapid vs. late vs. no relapse. Signatures visualized as relative values (Z-score) with rapid relapse (red), late relapse (green), and no relapse (blue).

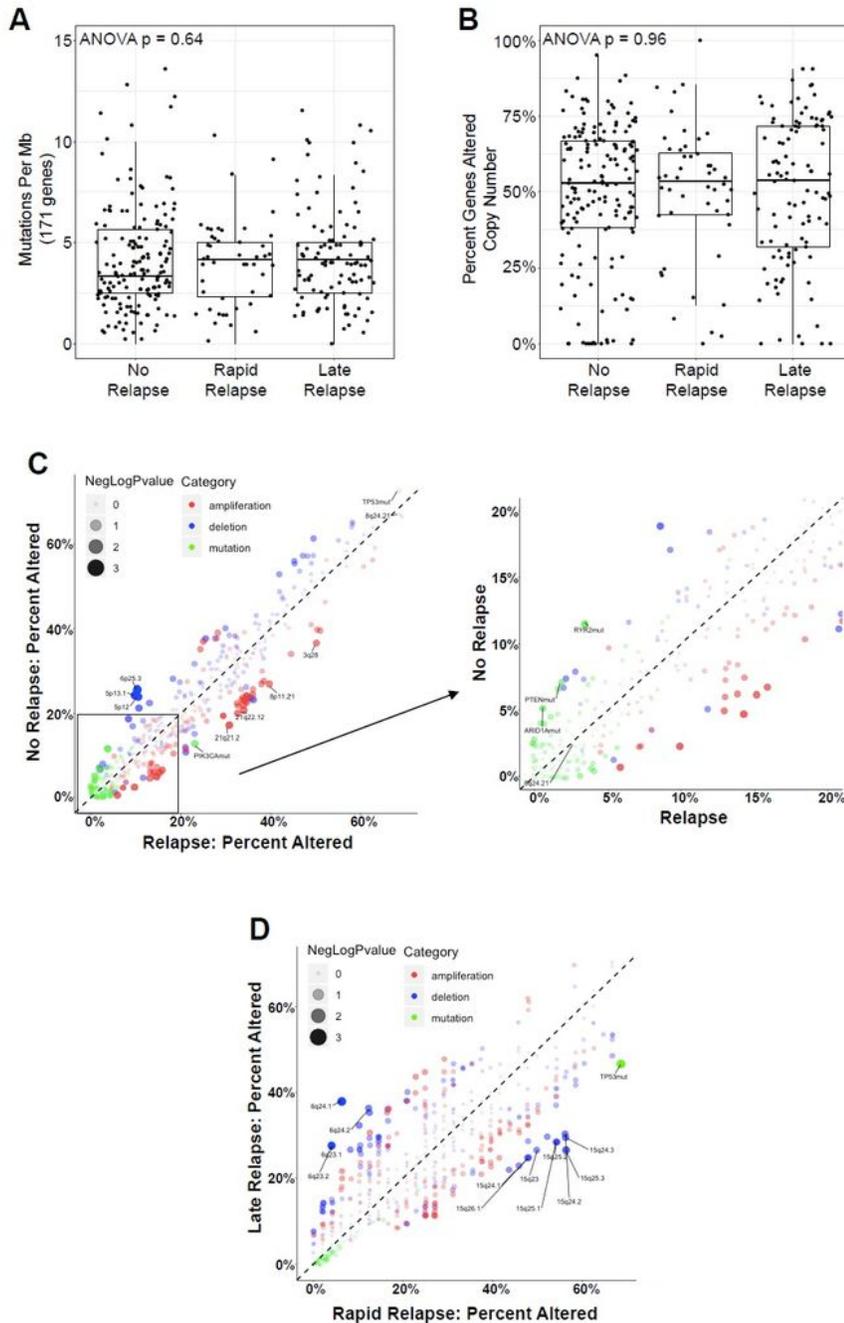


Figure 4

Mutations and copy number alterations in rapid vs. late vs. no relapse TNBCs. (A) Mutations per megabase of 171 cancer-related genes. (B) Percent genes altered by copy number gain (GISTIC 1 or 2) or loss (GISTIC -1 or -2). (C) Frequency of alteration of 171 cancer-related genes (green dots), copy number gains (red dots) or losses (red dots) by cytoband among rapid relapse (x-axis) vs. no relapse (y-axis) TNBCs (C) or rapid relapse (x-axis) vs. late relapse (y-axis) TNBCS (D). Size of dot indicates negative log of p-value for Fisher exact test with those genes and cytobands indicated demonstrate nominal $p < 0.05$. Zoomed-in image of those alterations with $< 20\%$ frequency indicated in right panel.

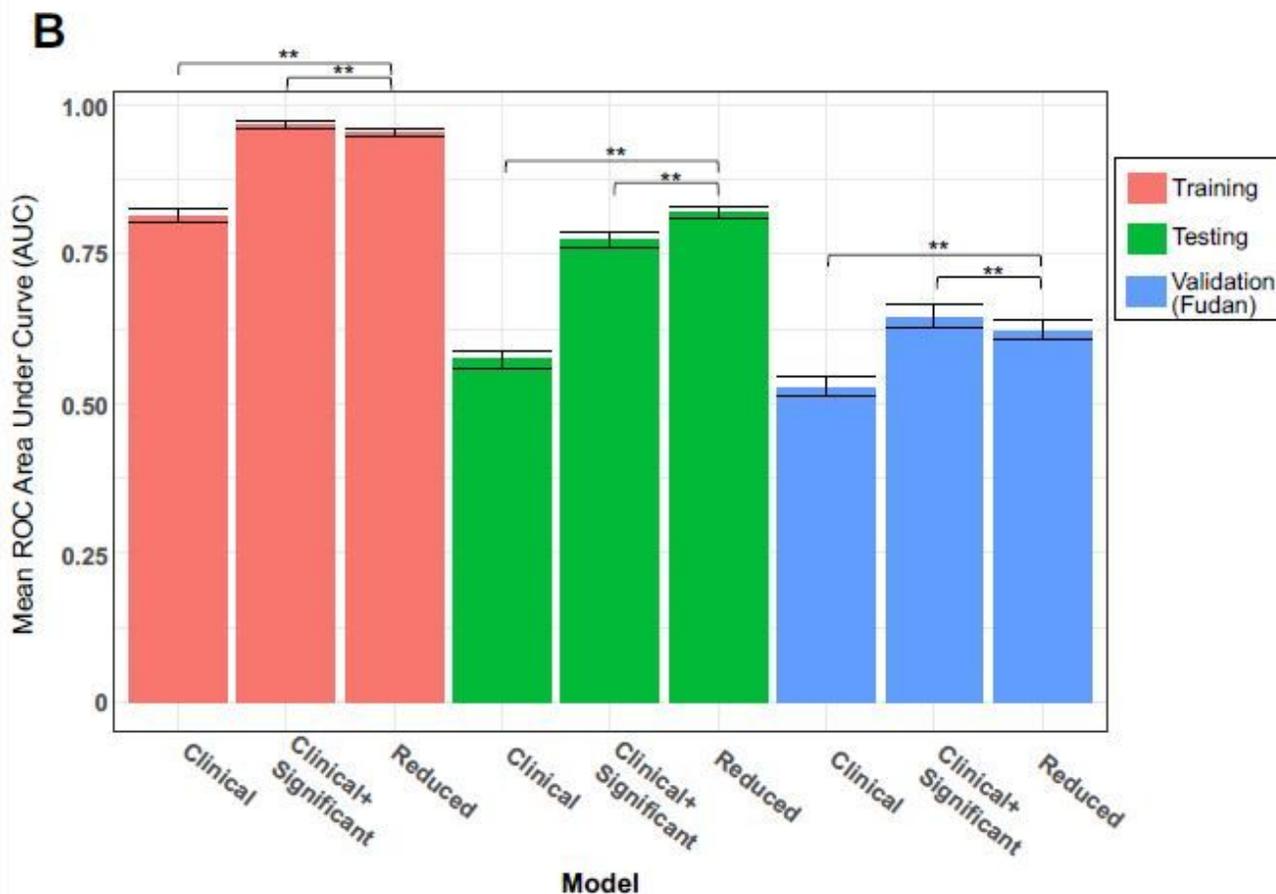
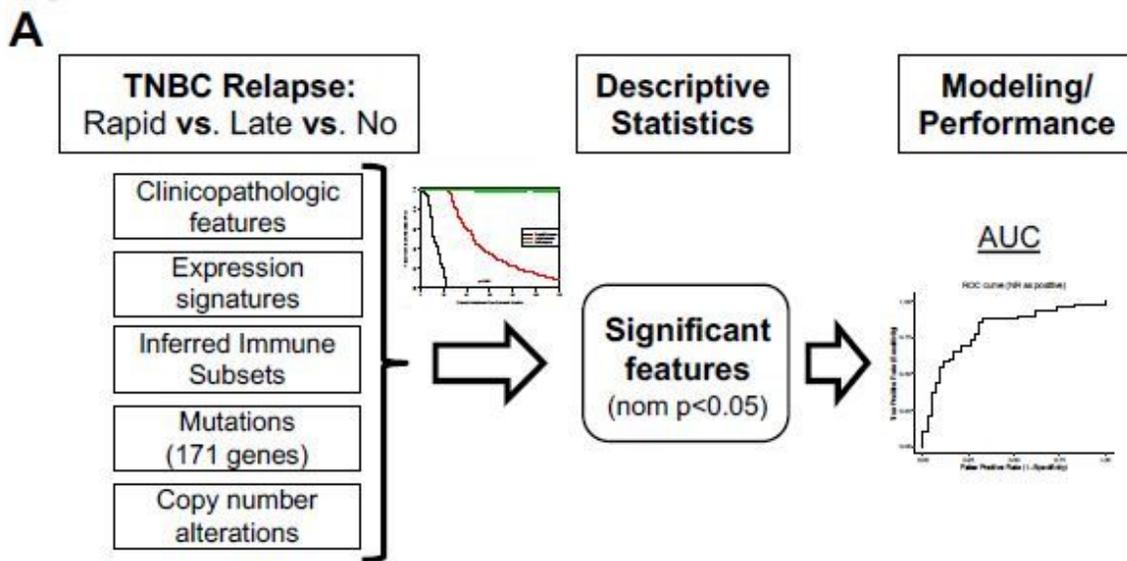


Figure 5

Developing an optimal clinical and multi-omic model of rapid vs. late relapse in TNBC. (A) Schematic of experimental steps including definition of variables, descriptive statistics, comparative modeling including model tuning, and assessment of model performance. (B) Receiver-operator characteristic (ROC) plots for each model's performance, measured by average area under the curve (AUC) of 25 independent runs of the train-test split. Each model was tuned to ensure optimal performance. Models are grouped and colored by cohort—red indicates training data (n=63), green indicates testing data (n=63), and blue indicates the independent validation Fudan cohort (n=34). For each grouping, the three models shown are: 1) “null model”, including only clinical variables; 2) “null plus significant features”, adding any feature significantly different between rrTNBC and lrTNBC with a nominal p-value < 0.05; and 3) “null plus significant features reduced”, including only features from model 2 that are among the top 25 most important genes in at least half of the independent runs. Asterisks indicate significance by Wilcoxon rank sum, * indicates $p < 0.05$, ** indicates $p < 0.01$, NS indicates “not significant” ($p > 0.05$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.pdf](#)