Proteomics Landscape of Triple-Negative Apocrine Breast Carcinoma Reveals Molecular Mechanisms of Tumorigenesis and Characteristics of Chemo-insensitivity

Yiying Zhu[1,5*], Mengping Long[2,5], Wenhao Shi[1,5], Tianlong He[3,5], Fangzhou Xie[3], Annan Qian[3], Yuqiao Liu[3], Taobo Hu[4*], Shaojun Tang[3*]

[1] Department of Chemistry, Tsinghua University, Beijing 100084 China. [2] Department of Breast Surgery, Peking University People's Hospital, Beijing 100044 China. [3] Bioscience and Biomedical Engineering Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453 China. [4] Department of Pathology, Peking University Cancer Hospital, Beijing, 100083 China. [5] These authors contributed equally: Y. Z., M. L., W.S., T.He. *Corresponding author email: yiying_zhu@mail.tsinghua.edu.cn (Y.Z.), shaojuntang@ust.hk (S.T.), thuac@bjmu.edu.cn (T.Hu.)

ABSTRACT

Triple-negative apocrine breast carcinoma (TNAC) is a rare type of triple-negative breast cancer (TNBC) characterized by apocrine morphology. Due to its low clinical incidence, TNAC has not been well studied at the molecular level in terms of cellular mechanisms of tumorigenesis and treatment response. In this study, we collected formalin-fixed paraffin-embedded samples from 31 patients diagnosed with invasive TNAC. Mass spectrometry analysis and whole-exome sequencing were performed to systematically construct the proteome and mutation landscape of the TNAC. Comparing ductal carcinoma in situ (DCIS), invasive, and adjacent normal tissues, we observed increased cadherin binding and ligase activity during tumorigenesis, heightened extracellular matrix signaling at tumor initiation, and elevated GTPase activity during tumor progression. Besides, we discovered genes such as C3 and COL18A1 enriched in somatic mutations are also significantly dysregulated in invasive TNAC. Additionally, by assessing biopsies before neo-adjacent chemotherapy, we provided molecular evidence supporting clinical observations of chemotherapy insensitivity in TNAC patients. We further identified a protein panel (CAPN1, CORO1B, HK1, RAP1B, and VPS4B) differentiating TNAC from non-apocrine TNBC, which could potentially serve as diagnostic markers. Taken together, this study represents the first large-scale proteomics analysis of TNAC and may provide guidance on decisions regarding the chemotherapy treatment of TNAC patients.

INTRODUCTION

Apocrine carcinoma (AC) is a unique type of breast cancer characterized by apocrine morphology, which includes large tumor cell size, rich eosinophilic granules, and a significant nucleolus[1, 2]. It only accounts for less than 1% of all breast cancers[3]. AC is mostly ductal, either in situ (known as DCIS) or invasive, meaning cancer cells grow within the ducts or invade nearby breast tissues after penetrating the basal membrane. Invasive apocrine carcinoma, according to the WHO, is a tumor characterized by over 90% of its cells having apocrine morphology[4]. Almost half of AC is triple-negative (TN), known as triple-negative apocrine carcinoma (TNAC), lacking estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), thus cannot be effectively treated by drugs targeting ER/PR/HER2. Additionally, TNAC tumor tissues typically express low Ki67 and strong androgen receptor (AR), indicating a low proliferation rate and potential for treatment using AR-targeted drugs[5]. TNAC usually responds poorly to chemotherapy[6-8], but has a better prognosis than non-apocrine triple-negative breast cancer (NA-TNBC)[5, 9-12]. Due to the low clinical incidence, there is limited research on characterizing TNAC using both genomic and proteomic sequencing data. Previous research primarily focused on genomic profiling to determine subtypes and potential drug targets for TNAC[13, 14]. There remains a lack of proteomics study to assess the dysregulated proteins in TNAC and changes in the proteomic landscape following neo-adjacent chemotherapy.

To date, multidrug chemotherapy is commonly used in the treatment of malignant tumors, while neoadjuvant chemotherapy, the chemo-treatment administered before surgery to shrink the tumor, is gaining popularity[15-17]. Due to its side effects, there is a growing interest in personalized treatment by identifying patients who may not respond to chemotherapy[18, 19]. Genetic tests combined with endocrine therapy have been successful in treating luminal type (ER/PR+) breast cancer patients. For instance, studies like TAILORx, RxPONDER, and MINDACT have utilized gene tests to determine recurrence risk scores for ER+/PR+/HER2- patients for chemo-exemption[20-22]. There has also been research on replacement therapies for HER2+ cancer, such as HER2-targeted drug trastuzumab[23]. Regarding chemo-exemption for TNBC patients, the development of immunotherapy, including pembrolizumab, and new drug targets such as AR, offers potential alternatives[24-27]. Increasing evidence suggests that TNAC patients might not

benefit from chemotherapy, leading to discussions about identifying molecular biomarkers to screen for TNAC patients to exclude from chemotherapy[6-8].

In this study, we aim to systematically characterize aberrantly expressed proteins, perturbated pathways, and heavily mutated genes in TNAC patients. We mainly performed mass spectrometry proteomics profiling to construct the proteome landscape of the tumor and matched normal samples. This helped us identify significantly changed proteins in tumor tissues and decipher potential molecular features and mechanisms of tumor initiation and progression. We then applied whole-exome sequencing (WES) to characterize somatically mutated genes and uncover their link to aberrantly expressed proteins. Additionally, we investigated the treatment response of chemotherapy based on proteome profiles of TNAC tumor tissues, which provided the basis for us to evaluate the insensitivity of the chemotherapy at the molecular level. Finally, we compared TNAC and ductal NA-TNBC and proposed a potential TNAC molecular diagnostic marker panel based on five proteins using machine learning techniques. Overall, this study provided the first-of-its-kind proteomics landscape and served as invaluable molecular research for potential precision medicine in TNAC treatment.

METHODS

Tissue Samples

All of the cases diagnosed as invasive breast apocrine carcinoma between 2008 and 2021 in Peking University Hospital were retrieved and reviewed by two senior pathologists. The diagnostic criteria include a minimum of 95% of the tumor displaying apocrine differentiation, tumor cells with a minimum Nuclear:Cytoplasm ratio of 1:2, abundant eosinophilic cytoplasm, distinct nucleoli, and well-defined cell boundaries. Patients with positive expression in ER or PR were excluded. Patients with HER2-amplification were also excluded. A total of thirty-one cases of invasive TNAC were included for proteomic analysis. The Formalin-fixed paraffin-embedded (FFPE) blocks used in this study were obtained from Peking University Cancer Hospital in Beijing, China, with approval from the hospital's ethical committee under reference number 2020KT113. The cells were stained using Hematoxylin and Eosin (H&E), and the expression of protein markers such as ER, PR, and HER2 was evaluated using immunohistochemical staining (IHC), as described in a

previous publication[7]. Out of the 31 invasive TNAC cases, 10 still contained accompanying DCIS lesions; therefore, invasive, DCIS, and adjacent normal tissues (~10 μm thick) were carefully dissected from the FFPE blocks on a macroscopic level. In addition, seven patients had received neoadjuvant therapy, and FFPE samples of their needle biopsies taken before chemotherapy were included in the study. The survival of these patients was evaluated using the Miller–Payne grading system.

Sample preparation for proteomics analysis

Paraffin of FFPE samples was solubilized by xylene, and protein pellets were washed with ethanol. The ethanol was removed completely and the sections were left to air-dry. Samples were lysed with a buffer containing 4% sodium dodecyl sulfate (SDS), 0.1 M Tris-HCl (pH 8.0), 0.1 M DTT (Sigma, 43815), and 1 mM PMSF (Amresco, M145). The collected solution was mixed with pre-cold acetone in a 4:1 ratio. After precipitating the proteins with acetone, they were further washed with cooled acetone. The protein precipitation was then redissolved in 200 μL of 8 M urea solution. The protein concentration was determined by the bicinchoninic acid (BCA) protein assay (Thermo Scientific, 23227). Samples containing 100 μg of proteins were first reduced with 10 mM dithiothreitol at 56 °C for 30 min and alkylated with 10 mM iodoacetamide at room temperature (RT) in the dark for an additional 30 min. Then the samples were digested by trypsin using the filter-aided proteome preparation (FASP) method. Specifically, the samples were transferred into a 30 kD Microcon filter (Millipore, MRCF0R030) and centrifuged at 14,000 × g for 20 min. The precipitate on the filter was washed twice by adding 300 μL of washing buffer (8 M urea in 100 mM Tris, pH 8.0) into the filter and then centrifuged again at 14,000 × g for 20 min. The precipitate was then resuspended in 200 μL of 100 mM NH_4HCO_3. Trypsin was added to the filter at a protein-to-enzyme ratio of 50:1 (w/w), and the proteins were digested at 37 °C for 16 h. After this, peptides were collected by centrifugation at 14,000 × g for 20 min and dried in a vacuum concentrator (Thermo Scientific).

LC-MS/MS analysis

The dried peptide samples were re-dissolved in Solvent A (0.1% formic acid in water) and loaded onto a trap column (100 μm × 2 cm, home-made; particle size, 3 μm; pore size, 120 Å; SunChrom, USA) under a maximum pressure of 280 bar using Solvent A. The samples were then separated

on a home-made 150 μm × 12 cm silica microcolumn (particle size, 1.9 μm; pore size, 120 Å; SunChrom, USA) with a gradient of 5-35% mobile phase B (acetonitrile and 0.1% formic acid) at a flow rate of 300 nL/min for 120 min.

The eluted peptides were ionized under 2.2 kV. Mass Spectrometry (MS) operated under a data-dependent acquisition (DDA) mode. For detection with the Orbitrap Eclipse mass spectrometer, a precursor scan was performed in the Orbitrap by scanning m/z 300-1,500 with a resolution of 60,000. Then, MS/MS scanning was carried out in the Orbitrap by scanning m/z 200-1,400 with a resolution of 15,000. The most intense ions selected under top-speed mode were isolated in Quadrupole with a 1.6 m/z window and fragmented by higher energy collisional dissociation (HCD) with a normalized collision energy of 32%. The maximum injection time was set to 40 ms for full scans and 30 ms for MS/MS scans. Dynamic exclusion time was set as 30 s.

MS data processing

All MS data were processed in the Maxquant（v 1.6.17.0）platform. The raw files were searched against the human National Center for Biotechnology Information (NCBI) ref-seq protein database (updated on 07-04-2013, 32,015 entries). Mass tolerances were set at 10 ppm for precursor ions and 0.05 Da for product ions. Up to two missed cleavages were allowed. Carbamidomethylation at cysteine was specified as a fixed modification, while acetylation at protein N-terminus and oxidation at methionine were considered as variable modifications. The data were also searched against a target-decoy database to ensure identifications were accepted at a false discovery rate (FDR) of 1% at both peptide and protein levels. Proteins were classified into protein groups when identified peptides could not differentiate protein isoforms and homologues. The Maxquant-iBAQ labeling-free quantitation method was applied to evaluate protein abundances. Missing values were imputed using minimum values in the corresponding sample. An R package "DEP" (v 1.16.0) was used for data preprocessing and manual examination of data. Prior to processing, the proteins were filtered to ensure their presence in the majority of the samples and the number of missing values fell in the allowed range. The filtered datasets were than log2-transformed and normalized by variance stabilizing transformation to eliminate the effect of different variance among proteins in DEP. The *removeBatchEffect* function in "limma" package was utilized to remove batch effect between TNAC and NA-TNBC datasets. The *ComBat* function in "sva" package (v 3.42.0) was

also employed for comparison.

Gene Ontology (GO) and KEGG enrichment analysis

Differential protein expressions were analyzed by the R package "limma" (v 3.50.0). Specifically, a separate linear model was fit for each protein, and contrasts were then tested. Empirical Bayes moderation is carried out to obtain a more precise estimate of protein-wise variability. Generally, proteins with an adjusted p-value < 0.05 and a |log2 Fold change| > 1 were identified as differentially expressed proteins. Additionally, R package "clusterProfiler" (v 4.2.2) with default parameters was applied. For the GO enrichment analysis, we focused on enriching "BP" (biological process) and "MF" (molecular function) subontology terms.

Gene set enrichment analysis (GSEA) and Gene set variation analysis (GSVA)

For GSEA analysis, R package "clusterProfiler" (v 4.2.2) with default parameters was utilized by the following annotation gene sets from MSigDB (v 2023.2.Hs): H (hallmark gene sets), CP (canonical pathways gene sets) and C5 (ontology-gene sets). Differential expression analysis was conducted on gene set expression scores using R package "limma". We conducted a single sample GSEA (ssGSEA) using R package "GSVA" (v 1.42.0). Specifically, the KEGG_MEDICUS subset of CP gene sets was evaluated. To reduce the redundancy of highly similar gene sets, we computed the overlap among gene sets using the function "Compute Gene Sets Overlap". For gene sets with a similarity value larger than 0.8, only one gene set was randomly chosen to be preserved. Similarly, we deployed "limma" to perform differential analysis between TNAC and NA-TNBC tumor tissues. Pathways with an adjusted P value < 0.01 were identified as significantly altered pathways.

Weighted correlation network analysis (WGCNA)

WGCNA analysis was performed to identify protein modules during TNAC tumor development using R package "WGCNA" (v 1.72-5). 3277 proteins that passed filtering were used for co-expression network construction. Power 10 was selected as the soft threshold as it fits a scale-free topology model well (R2=0.87) while preserving a connectivity of 40.72. A signed network was constructed to keep track of the sign of the co-expression information. A topological overlap measure (TOM) was used to evaluate the network interconnectedness and protein modules composed of highly connected proteins were then identified by hierarchical clustering. The

modules identified using the blockwiseModules function were further annotated using the compareCluster function in the R package "clusterProfiler." This was based on annotation gene sets from MSigDB, including H (hallmark gene sets), CP (canonical pathways gene sets), and C5 (ontology-gene sets). The top significantly enriched terms (adjusted $p < 0.01$) were manually selected to summarize the functions of the modules.

WES

WES was conducted on six paired invasive and adjacent normal tissue samples from three TNAC patients. The sequencing results were aligned to the hg19 reference genome using BWA (v 0.7.8-r455), and the depth was checked by Sambamba (v 0.6.8). Aligned BAM files were processed by SAMtools (v 1.9) to identify the SNP sites and INDEL sites. The SNP sites were annotated using ANNOVAR (v 2017June8) with reference databases, including Refseq, dbSNP, 1000 genome, GO, and KEGG. The final annotations were consolidated into MAF format. Subsequent analyses were conducted based on the samples' origin using maftools and manual scripts. No threshold for statistical significance was applied.

Lasso regression

The *glmnet* function from R package "glmnet" (v 4.1-3) was leveraged to perform Lasso regression. Significantly up/down-regulated proteins (selected as described above) that were uniquely detected in TNAC or NA-TNBC were selected for model establishment. Tumor samples from TNAC and NA-TNBC patients were assigned labels "1" and "0", respectively. By setting the "family" argument to "binomial", we invoked a logistic regression model to solve sample labels using the expression matrix. The "alpha" argument was set to 1 to apply the lasso penalty in model training. We first fit a Lasso regression model using the full data, then performed a 10-fold cross-validation to determine the best lamda parameter. Finally the model with the lowest misclassification error was selected when lambda = 0.105. In the final model, five variables were retained with a non-zero coefficient. To address the possibility of overfitting due to the relatively small sample size, we constructed models using the same protein expression matrix and randomly assigned sample labels. After 1000 repeats, the random models failed to achieve a comparable performance as the real data.

RESULTS

Clinical and pathological characteristics of samples

In order to explore TNAC as a unique tumor type, we collected formalin-fixed paraffin-embedded (FFPE) samples from 31 patients diagnosed with invasive TNAC (Table S1). The median age of patients at diagnosis was 57 years, ranging from 38 to 79. 41.9% (13/31) of patients were at TNM stage I, 45.2% (14/31) were at stage II, and 12.9% (4/31) were at stage III. The morphology of samples was evaluated and confirmed by two pathologists, and protein markers were confirmed by IHC. Example HE and IHC images are shown in Figure 1A/B. All patients tested negative for the hormone receptors ER and PR, and were non-amplified for HER2. They had positive AR and low Ki67, along with an increasing Epidermal growth factor receptor (EGFR) from normal to invasive tissues. The average duration of storage for FFPE samples was 1869 days, ranging from 229 to 3185 days. Ten patient samples still contain DCIS tumor tissues. Seven patients received neoadjuvant chemotherapy before surgery, and their needle biopsy FFPEs were also collected. Among 7 neoadjuvant chemotherapy patients, 3 were at TNM stage II, 2 were at stage III, and 1 was at stage I, with a median age of 54 (from 46 to 63). Pathological responses to neoadjuvant chemotherapy were assessed using the Miller–Payne (MP) grading system on a five-point scale. Four patients had an MP grade 2 response, while 3 had a grade 3 response. In conclusion, a total of 79 samples were collected in the study, including 31 adjacent normal tissues, 10 DCIS tissues, and 38 invasive tissues (from 31 patients, including 7 pre-chemotherapy biopsies). Invasive and normal tissues from three patients were further analyzed by WES. The study also includes previously collected proteomics data from fresh-frozen breast tissue samples of ductal invasive NA-TNBC patients[28]. These NA-TNBC patients had a median age of 57.5 years (ranging from 41 to 77). Half (3/6) were at TNM stage II, while 2 were at stage III, and 1 was at stage I.

TNAC tumor initiation and progression

Breast cancer is believed to arise from the epithelial cells within the terminal ductal lobular units (TDLU). Prior to the breaching of the basement membrane, it is denoted as DCIS, which is traditionally regarded as lacking the capacity for distant metastasis. Upon penetration of the

basement membrane and subsequent invasion into the peritumoral stroma, the disease is classified as invasive breast cancer, characterized by its potential for distant metastasis. Consequently, the therapeutic approach necessitates the implementation of systemic therapies to prevent or manage metastatic spread. Understanding their molecular characteristics can help us grasp the initiation and progression of the tumor. Invasive and DCIS tissues, along with adjacent normal tissues, were carefully removed from selected FFPE samples, and underwent mass spectrometry proteomics analysis. In total, 5952 protein groups were quantified in this dataset, with 4065 proteins from normal tissues, 3844 proteins from DCIS tissues, and 5277 proteins from invasive tissues.

Principal component analysis (PCA) was performed on the proteomic data from three different tissue types. As shown in Figure 2A, the components clustered distinctly, with clear boundaries between normal, DCIS, and invasive tissues, indicating significant molecular differences among them. This result demonstrates that proteomics can effectively reflect the molecular distinctions between different tumor tissues. We then identified differentially expressed proteins using the linear model implemented in "limma", screening a total of 2309 proteins with significant differential expression across the three tissue types. Bilateral clustering of these differentially expressed proteins (Figure 2B) showed that each tissue type formed distinct clusters, highlighting significant differences in protein expression. Additionally, it was observed that DCIS and invasive tissues were more similar to each other, with many proteins expressed in DCIS tissues being lost or down-regulated in invasive tissues. This suggests that studying the functions and signaling pathways of these differentially expressed proteins may uncover the molecular mechanisms underlying TNAC.

We utilized WGCNA to investigate the co-expression network of differentially expressed proteins (Figure 2C). Our analysis grouped 3277 proteins into five modules (ME0-ME4). The ME0 module included proteins associated with actin binding. Proteins related to cell localization and regulation, including those involved in Golgi vesicle transport, were found in the ME1 module. ME2 was linked to MYC target proteins and oxidative phosphorylation. The ME3 module was associated with the extracellular matrix and epithelial-mesenchymal transition. ME4 contained proteins related to chromatin DNA binding. Together, these modules reflected the proliferative and metastatic characteristics of tumor cells. Additionally, we performed bilateral clustering of the five modules and observed distinct tissue-specific significance (Figure 2D). ME3 was most prominent

in normal tissues, ME0 in DCIS tumor tissues, and ME1 in invasive tumor tissues. These findings are consistent with established pathological knowledge.

Under detailed data analysis, we found 1647 upregulated proteins and 463 downregulated proteins in invasive tumors. The GO enrichment analyses (Figure 3) revealed that the upregulated proteins in invasive tumors were primarily associated with cadherin binding and GTP binding, while the downregulated proteins were mainly involved in the extracellular matrix and actin binding. Next, we compared DCIS tumor tissue with normal tissue and identified 812 upregulated proteins and 458 downregulated proteins. In DCIS tumors, the upregulated proteins were also associated with cadherin binding and ligase activity, while the downregulated proteins were concentrated in the extracellular matrix and actin binding. These findings indicate that DCIS and invasive tumors exhibit similar protein expression changes. Finally, we compared the proteomic data between invasive and DCIS tumor tissues, identifying 304 highly expressed proteins in invasive tumors and 101 highly expressed proteins in DCIS tumors. Invasive tumor tissues showed enrichment in proteins associated with GTPase activity and guanyl ribonucleotide binding signaling pathways. In contrast, DCIS tumor tissues were enriched in proteins related to the extracellular matrix and sulfur compound binding signaling pathways. The GSEA results (Figure S1) revealed that the mitochondrial translation process was enhanced in invasive tissues, while cholesterol synthesis was dysregulated in DCIS tissues. Both types of tumor tissue showed inhibition in the complement signaling pathway. Additionally, when comparing DCIS and invasive tissues, we observed stronger epigenetic regulation of gene expression in invasive tissues and increased activity in the scavenging by class A receptor signaling pathway in DCIS tissues.

WES of TNAC

WES analysis was conducted on six paired samples of invasive tumor tissues and normal tissues. Figure 4 provides an overview of the WES sequencing results. The differential analysis yielded 32 mutated genes. Splice sites were the dominant mutation type, followed by in-frame insertions and deletions (Figure 4A). The most common single nucleotide variation (SNV) class was C to T transitions (Figure 4B). Through genomic sequencing, top ten mutated genes, including DSPP and NRDC, were identified as shown in Figure 4C.

We further compared these loci with proteomic results and identified twelve overlapping loci and their corresponding proteins: ACE2, C3, COL18A1, DNM3, HCLS1, MAL2, NRDC, NUP54, OLFM4, PRKCSH, SMARCA2, TNS1 (Figure 4D). Among these genes, C3 exhibited the most significant changes in protein expression, suggesting that its gene mutation may have cis-effects on the protein and play crucial roles in the regulation of TNAC. Detailed analysis of the mutations in the overlapping genes revealed that C3, ACE2, MAL2, OLFM4, and DNM3 had splice site mutations; TNS1, HCLS1, and NUP54 had nonsense mutations; COL18A1 had frameshift deletions and NRDC, PRKCSH and SMACR2 have in-frame deletion. Example mutation sites are shown in Figure 4E.

Complement C3 has been reported to be a key molecule in the activation of B lymphocytes[29]. Therefore, the mutation of C3 gene and the reduction of C3 protein level may indicate that there is a disorder in B cell activation and immune suppression in this group of patients. Similarly, the gene COL18A1, which has been reported to be significantly upregulated in human breast cancer and strongly associated with poor prognosis in high-grade breast cancer, is also mutated, and its protein expression is reduced. Ablation of COL18A1 in breast cancer significantly improves the efficacy of HER2-targeted therapy[30]. In summary, WES results identified key mutated genes and integrated analysis with protein expression differences, providing potential mechanism clues for the occurrence and development of breast cancer from the perspectives of cellular, extracellular, and cellular immunity.

The impact of chemotherapy on TNAC

Based on our clinical observations and database analysis, we have found that TNAC patients generally do not respond well to chemotherapy, but they do have a higher survival rate than other types of TNBC patients[7]. This raises questions about the necessity of administering chemotherapy to TNAC patients. In order to explore this further, we examined paired samples from 7 TNAC patients taken before and after chemotherapy treatment. We set to compare the proteomes of these samples and investigate any changes in protein expression within the tumors.

PCA analysis revealed that tumor tissues before and after chemotherapy treatment had significant differences compared to normal tissues. Tumor tissues showed more remarkable similarity to each other before and after treatment, as indicated in Figure 5A, consistent with the clinical observation of the poor pathological response to chemotherapy. We further analyzed the differentially expressed proteins in pre- and post-treatment samples. A total of 62 proteins were upregulated, and 103 proteins were downregulated, as shown in Figure 5B. For instance, proteins such as VPS8 were highly expressed only in pre-treatment tumor tissues, while proteins like ORM1 were highly expressed in normal tissues and showed increased expression after treatment. Additionally, proteins such as FUCA1 were upregulated only in post-treatment tumor tissues. We visualized the limited number of significantly differentially expressed proteins using a heatmap (Figure 5C). Protein-protein interaction networks on down-regulated proteins don't show any significant correlations, but the networks on up-regulated proteins show that a few proteins are involved in the acute inflammatory response (Figure 5D). The existing literature indicates that chemotherapy may induce acute inflammatory reactions[32, 33].

Comparison of TNAC and NA-TNBC

TNAC and NA-TNBC subtypes exhibit significant differences in clinical prognosis and treatment response. In this study, we compared the TNAC proteome data with published ductal invasive NA-TNBC proteome data to differentiate these two cancer types at the molecular level. First, we identified differentially expressed proteins in TNAC and NA-TNBC after normalization (Figure S3). In TNAC, we found 897 upregulated proteins and 312 downregulated proteins. In NA-TNBC, we identified 494 upregulated proteins and 237 downregulated proteins. Among these, 352 upregulated and 152 downregulated proteins were common to both TNAC and TNBC. The subsequent analysis focused on the 878 differentially expressed proteins that were unique to either TNAC or NA-TNBC.

We conducted a comparative analysis of proteomic data from normal tissues and invasive tissues from TNAC and NA-TNBC using GSVA (Figure 6A). We identified 38 signaling pathways exhibiting significant differences across the three tissue types ($p < 0.05$, $|\log2 \text{ Fold change}| > 1$). Specifically, 12 signaling pathways were enriched in normal tissues, 18 in NA-TNBC, and 35 in

TNAC. In NA-TNBC, pathways related to cytoskeletal regulation were notably enriched, including the RHOA signaling pathway and the CAMK calcium ion signaling pathway. Conversely, TNAC displayed enrichment in metabolically related signaling pathways, such as the tricarboxylic acid cycle, glycogen degradation, and the urea cycle. These findings suggest that NA-TNBC exhibits more pronounced alterations in cellular structure and performs a stronger propensity for metabolic dysregulation.

In addition to the morphology differentiation and traditional markers such as Ki67 to differentiate chemo-insensitive patients[18], we aim to discover new biomarkers that will help potential clinical diagnostics for chemotherapy exemption or de-escalation. We applied a Lasso logistic regression model to identify five protein markers that most effectively distinguish between TNAC and ductal NA-TNBC: CAPN1, CORO1B, HK1, RAP1B, and VPS4B (Figure 6C). Then, we evaluated the accuracy of this model. The model achieved an accuracy of approximately 1 with actual data and around 0.75 with randomly generated data (Figure 6B). Using these five protein markers, we calculated prediction scores for TNAC and NA-TNBC, finding that TNAC scores were significantly higher than NA-TNBC scores. Figure S3 illustrates the coefficients of the intercept and the five protein markers. All the single protein coefficients were negative, consistent with their expression profiles in TNAC and NA-TNBC; these proteins were upregulated in NA-TNBC but downregulated in TNAC.

DISCUSSION

TNAC is a rare type of TNBC indicated by unique morphology and protein markers. Previous studies have shown that TNAC typically expresses protein AR and EGFR while lacking ER, PR, and Ki67[2, 5, 7, 10, 11]. Current mass spectrometry proteomics data align with the protein expression observations from IHC (Figure 1 and Figure S4). Drawn by this TNAC proteomics data analyses, we found that ER, PR, and Ki67 were not detectable in tumor tissues, while EGFR was strongly expressed in invasive tissues, and AR was also present in invasive tissues. Notably, protein HER2 was expressed at a low level as observed in IHC images. Nevertheless, it is possible that current mass spectrometry fails to detect relatively low-abundant proteins due to its instrumentation[34]. For example, GCDFP-15 was not detected in our proteomics data but was detected in IHC (Figure S4).

Overall, the conclusion drawn from proteomics analyses is consistent with the IHC study and previous reports. It demonstrates the reliability of the data, showcasing the comprehensive protein characteristics of TNAC.

To date, this study is the first comprehensive proteogenomics profiling and molecular characterization of TNAC patient samples with chemotherapy. By comparing normal, DCIS, and invasive TNAC tissues, we observed both DCIS and invasive tissues had upregulated cadherin binding and ligase activity signaling pathways, however, significant differences were also found between DCIS and invasive tissues, indicated by elevated GTPase activity in invasive tissues whereas more prominent extracellular matrix signaling pathways in DCIS tissues. In addition, WES study revealed several heavily somatically mutation genes, including COL18A1, which had previously been found to be negatively associated with prognosis. Taken together, these findings augment our comprehension of the molecular mechanisms underlying TNAC and provide insights for future research on the regulation of molecular expression and pathways at the protein level.

Chemotherapy such as anthracyclines (e.g. doxorubicin), carboplatin (e.g. paraplatin), and taxanes (e.g. paclitaxel and docetaxel), usually causes DNA damage and inhibits mitosis[15]. However, we observed that there were no significant changes in the protein profiles of TNAC patients before and after chemotherapy, consistent with the observation that TNAC do not respond well to chemotherapy[7]. Although morphological differences between TNAC and NA-TNBC can be observed under a microscope, their differentiation still relies heavily on the expertise of pathologists. Currently, more than one pathologist is needed to confirm the diagnosis of TNAC. To assist in screening TNAC patients who may be suitable for chemotherapy exemption, we developed a potential marker panel to differentiate chemo-insensitive TNAC from chemo-sensitive NA-TNBC. As NA-TNBC samples were frozen fresh tissues, which are different compared to FFPE samples, we conducted a surrogate variable analysis to remove the batch effect to ensure consistent data across the normal tissues of TNAC and NA-TNBC datasets (Figure S3). We compiled a list of proteins that were expressed differently between NA-TNBC and TNAC. Using machine learning, we pinpointed five protein markers—CAPN1, CORO1B, HK1, RAP1B, and VPS4B—that exhibit the strongest statistical significance in distinguishing TNAC from NA-

TNBC. This set of potential biomarkers could help in distinguishing between the two conditions and could also aid healthcare providers in determining prognosis and making treatment decisions.

In conclusion, this study represents the first-of-its-kind comprehensive research on the TNAC proteomics landscape. The identified biomarkers and key regulatory pathways have the potential to play a crucial role in guiding clinical decisions regarding neo-adjuvant chemotherapy for TNAC patients.

REFERENCES

1. Thomas, A., Reis-Filho, J. S., Geyer, C. E., Jr., Wen, H. Y. Rare subtypes of triple negative breast cancer: Current understanding and future directions. *NPJ Breast Cancer* **9**, 55 (2023).
2. Vranic, S., et al. Apocrine carcinoma of the breast: A comprehensive review. *Histol Histopathol* **28**, 1393-1409 (2013).
3. Mills, A. M., et al. Pure apocrine carcinomas represent a clinicopathologically distinct androgen receptor-positive subset of triple-negative breast cancers. *Am J Surg Pathol* **40**, 1109-1116 (2016).
4. Tan, P. H., et al. The 2019 world health organization classification of tumours of the breast. *Histopathology* **77**, 181-185 (2020).
5. Saridakis, A., et al. Apocrine breast cancer: Unique features of a predominantly triple-negative breast cancer. *Ann Surg Oncol* **28**, 5610-5616 (2021).
6. Nagao, T., et al. The differences in the histological types of breast cancer and the response to neoadjuvant chemotherapy: The relationship between the outcome and the clinicopathological characteristics. In *Breast*, 2012/01/27 ed.; 2012; Vol. 21, pp 289-295.
7. Hu, T., et al. Triple-negative apocrine breast carcinoma has better prognosis despite poor response to neoadjuvant chemotherapy. *J Clin Med* **11**, (2022).
8. Montagna, E., et al. Prognosis of selected triple negative apocrine breast cancer patients who did not receive adjuvant chemotherapy. *Breast* **53**, 138-142 (2020).
9. Arciero, C. A., et al. Triple-negative apocrine carcinoma: A rare pathologic subtype with a better prognosis than other triple-negative breast cancers. *J Surg Oncol* **122**, 1232-1239 (2020).
10. Dellapasqua, S., et al. Immunohistochemically defined subtypes and outcome of apocrine breast cancer. *Clin Breast Cancer* **13**, 95-102 (2013).
11. Vranic, S., et al. Immunohistochemical and molecular profiling of histologically defined apocrine carcinomas of the breast. *Hum Pathol* **46**, 1350-1359 (2015).
12. Tsutsumi, Y. Apocrine carcinoma as triple-negative breast cancer: Novel definition of apocrine-type carcinoma as estrogen/progesterone receptor-negative and androgen receptor-positive invasive ductal carcinoma. *Jpn J Clin Oncol* **42**, 375-386 (2012).
13. Kim, J. Y., et al. Genomic characteristics of triple negative apocrine carcinoma: A comparison to triple negative breast cancer. *Exp Mol Med* **55**, 1451-1461 (2023).
14. Sun, X., et al. Invasive apocrine carcinoma of the breast: Clinicopathologic features and comprehensive genomic profiling of 18 pure triple-negative apocrine carcinomas. *Mod Pathol* **33**, 2473-2482 (2020).
15. Hassan, M. S.,Ansari, J.,Spooner, D.,Hussain, S. A. Chemotherapy for breast cancer (review). *Oncol Rep* **24**, 1121-1131 (2010).
16. Chiappa, C., et al. Neoadjuvant chemotherapy in breast cancer: Evaluation of the impact on surgical outcomes and prognosis. *Cancers (Basel)* **16**, (2024).

17. van den Ende, N. S., et al. Triple-negative breast cancer and predictive markers of response to neoadjuvant chemotherapy: A systematic review. *Int J Mol Sci* **24**, (2023).
18. Jiang, L., et al. Prospective observational study on biomarkers of response in pancreatic ductal adenocarcinoma. *Nat Med* **30**, 749-761 (2024).
19. Ulukaya, E.,Karakas, D.,Dimas, K. Tumor chemosensitivity assays are helpful for personalized cytotoxic treatments in cancer patients. *Medicina (Kaunas)* **57**, (2021).
20. Sparano, J. A., et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med* **379**, 111-121 (2018).
21. Kalinsky, K., et al. 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. *N Engl J Med* **385**, 2336-2347 (2021).
22. Piccart, M., et al. 70-gene signature as an aid for treatment decisions in early breast cancer: Updated results of the phase 3 randomised mindact trial with an exploratory analysis by age. *Lancet Oncol* **22**, 476-488 (2021).
23. Swain, S. M.,Shastry, M.,Hamilton, E. Targeting her2-positive breast cancer: Advances and future directions. *Nat Rev Drug Discov* **22**, 101-126 (2023).
24. Schmid, P., et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med* **382**, 810-821 (2020).
25. Traina, T. A., et al. Enzalutamide for the treatment of androgen receptor-expressing triple-negative breast cancer. *J Clin Oncol* **36**, 884-890 (2018).
26. Michmerhuizen, A. R.,Spratt, D. E.,Pierce, L. J.,Speers, C. W. Are we there yet? Understanding androgen receptor signaling in breast cancer. *NPJ Breast Cancer* **6**, 47 (2020).
27. Grellety, T.,MacGrogan, G.,Chakiba, C.,Kind, M.,Bonnefoi, H. Long-term complete response of an androgen receptor-positive triple-negative metastatic breast cancer to abiraterone acetate. *JCO Precis Oncol* **2**, 0 (2018).
28. Nan Wang, Y. Z., Lianshui Wang, Wenshuang Dai, Taobo Hu, Zhentao Song, Xia Li, Qi Zhang, Jianfei Ma, Qianghua Xia, Jin Li, Yiqiang Liu, Mengping Long, Zhiyong Ding. Parallel analyses by mass spectrometry (ms) and reverse phase protein array (rppa) reveal complementary proteomic profiles in triple-negative breast cancer (tnbc) patient tissues and cell cultures. bioRxiv
, 2024.
29. Lu, Y., et al. Complement signals determine opposite effects of b cells in chemotherapy-induced immunity. *Cell* **180**, 1081-1097 e1024 (2020).
30. Devarajan, R., et al. Targeting collagen xviii improves the efficiency of erbb inhibitors in breast cancer models. *J Clin Invest* **133**, (2023).
31. Xu, Y., et al. Eralpha is an rna-binding protein sustaining tumor cell survival and drug resistance. *Cell* **184**, 5215-5229 e5217 (2021).
32. Zhao, H., et al. Inflammation and tumor progression: Signaling pathways and targeted intervention. *Signal Transduct Target Ther* **6**, 263 (2021).
33. Feng, X., et al. Acute inflammatory response during neoadjuvant chemotherapy for locally advanced breast cancer: A case report. *Cureus* **9**, e1332 (2017).
34. Meissner, F.,Geddes-McAlister, J.,Mann, M.,Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat Rev Drug Discov* **21**, 637-654 (2022).

AUTHOR CONTRIBUTIONS

Study conception and supervision: Y.Z., M.L., S.T., T.Hu. Investigation and acquisition of clinical data (sample and patient information collection, pathology review): M.L., T.Hu. Investigation and acquisition of sequencing data: W.S., Y.Z. Data analysis, integration, and interpretation: T.He., F.X., A.Q., Y.L., Y.Z., S.T., W.S., T.Hu., M.L. Supervision of the bioinformatics analysis: S.T., Y.Z., T.Hu Writing: W.S., Y.Z., T.He., F.X. M.L. Reviewing and editing: Y.Z., S.T.

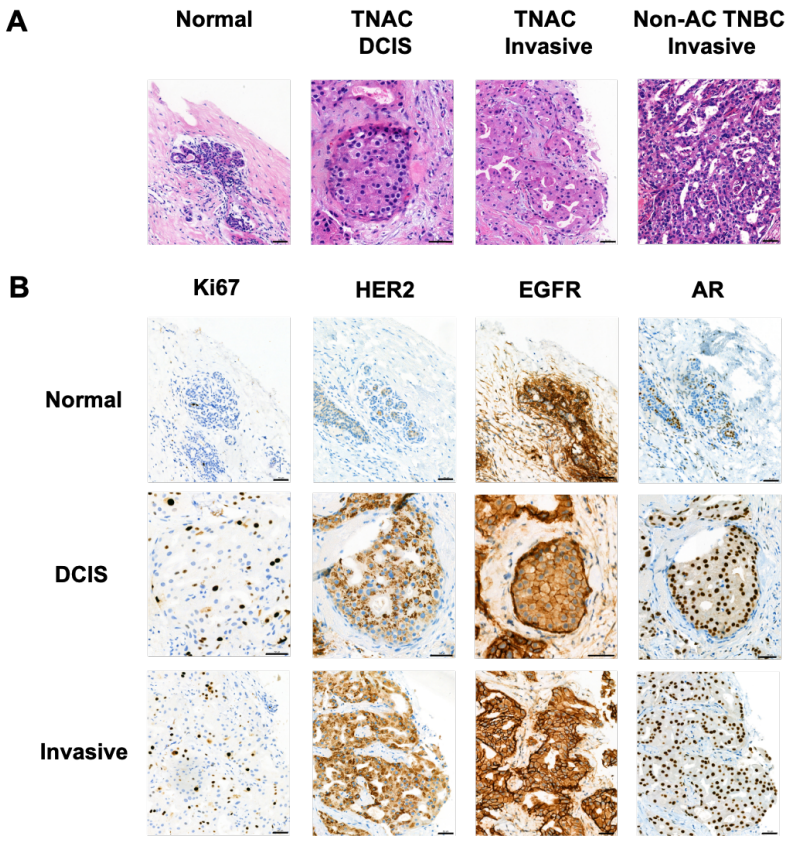

Figure 1. (A) Representative H&E-stained images from normal, TNAC DCIS, TNAC invasive, and NA-TNBC invasive tissues. These images show the unique apocrine morphology of TNAC tumors, characterized by abundant eosinophilic cytoplasm, distinct nucleoli, and well-defined cell boundaries. The scale bar indicates 50 μm. (B) Representative IHC images of Ki67, HER2, EGFR, and AR in normal, DCIS, and invasive tissues of a TNAC patient. The scale bar indicates 50 μm.

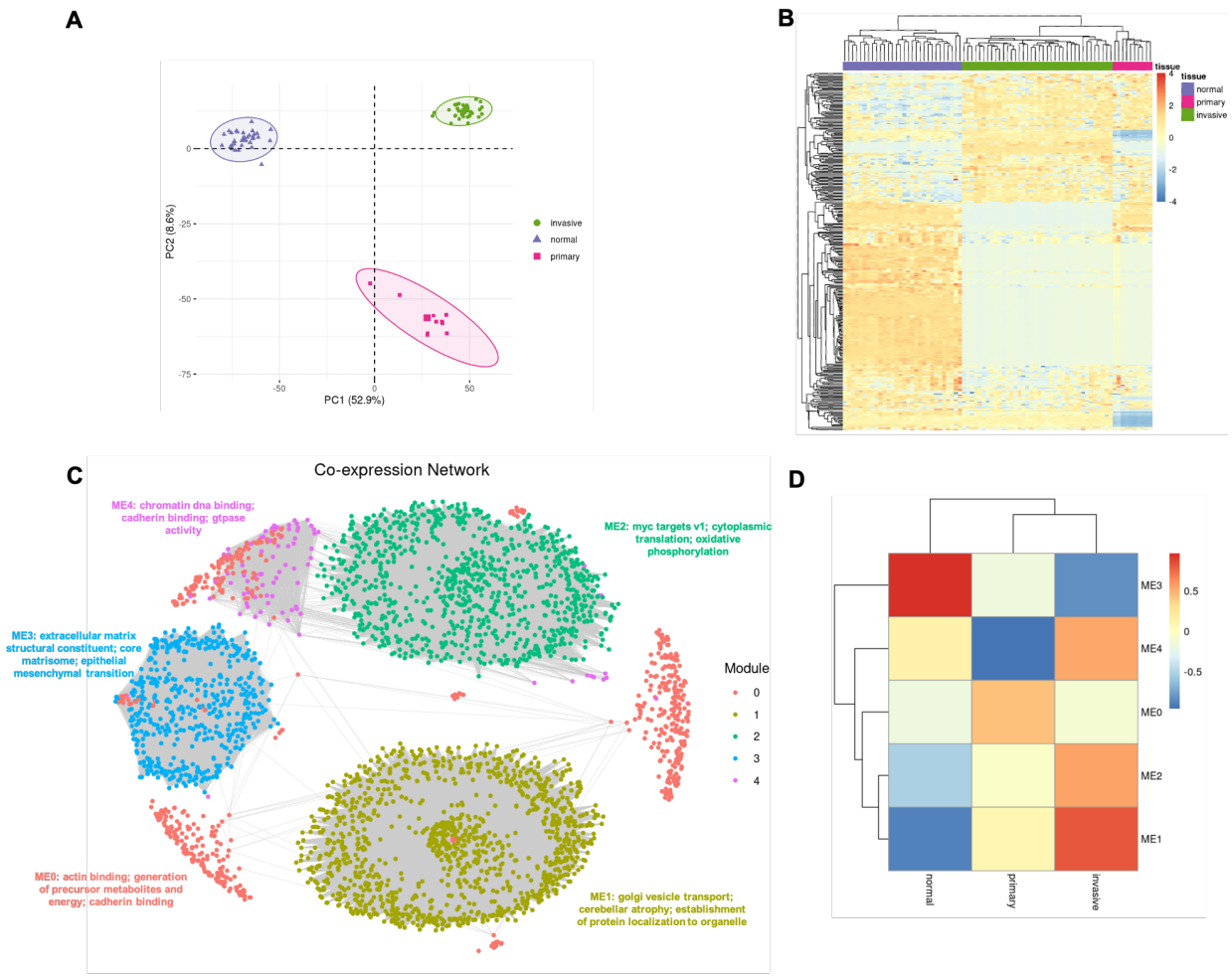

Figure 2. Proteomics landscape of normal, DCIS (primary), invasive TNAC tissues. (A) Principal component analysis (PCA) plot shows unsupervised clustering of normal, DCIS, and invasive tissues. Every dot on the graph symbolizes a sample, and the percentage value shows the amount of variance explained. (B) Heatmap depicts variations in protein expression, grouping normal, DCIS, and invasive tissues. Each cell illustrates the standardized score for protein abundance in a particular sample. The color scheme corresponds to the standardized score reflecting protein expression. (C) WGCNA protein co-expression analysis finds 5 functional protein modules (ME0–4) that are enriched in TNAC tumor samples. A network node represents each protein, color-coded based on the different functional modules. (D) Bilateral clustering of the five modules among normal, DCIS, and invasive tissues indicates that ME3 is most prominent in normal tissues, ME0 in primary tumor tissues, and ME1 in invasive tumor tissues.

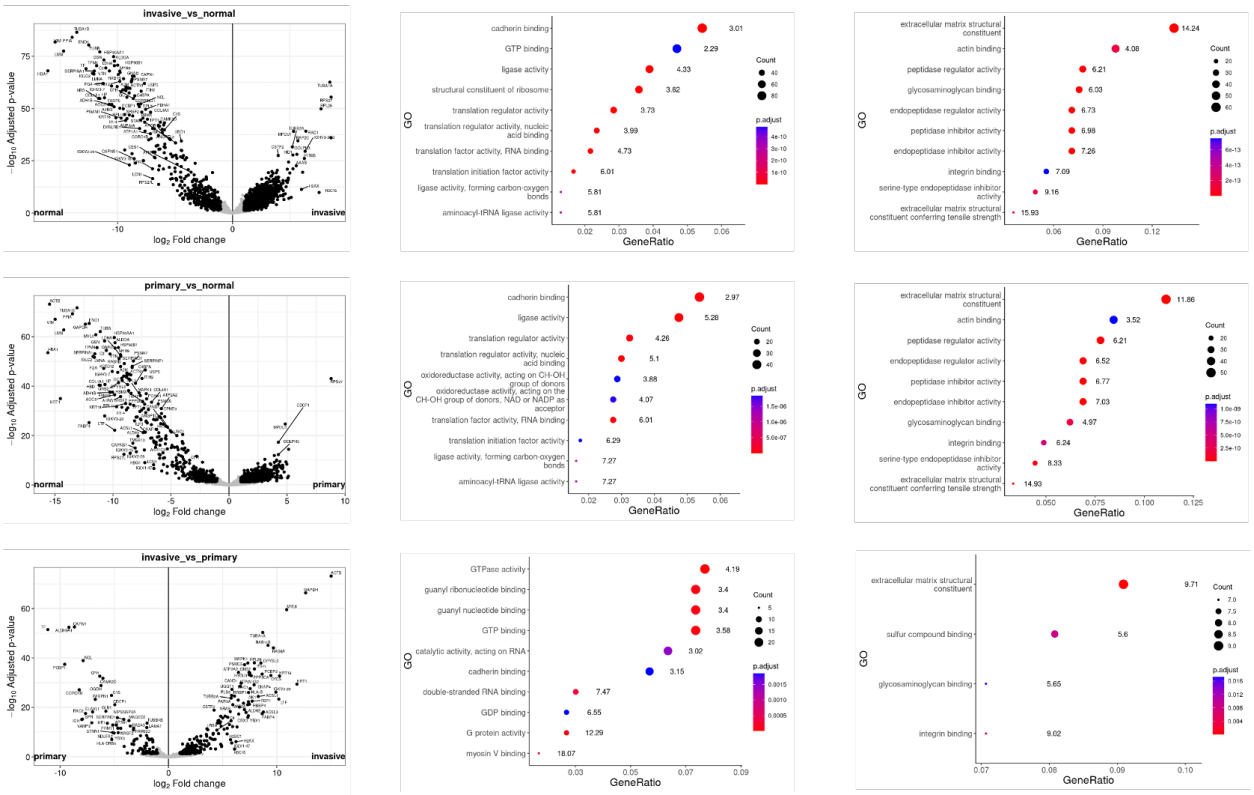

Figure 3: Differentially expression analysis of TNAC normal, DCIS (primary), invasive tissues. Left: Volcano plots showing differentially expressed proteins between three types of tissues: invasive vs. normal, DCIS vs. normal, and invasive vs. DCIS. The plots depict log2 Fold change versus –log10 p-value. The dark black dots are proteins defined as significantly changed ($p < 0.05$, |Log2 Fold change| > 1). Middle: Top enriched GO terms for upregulated proteins in each comparison based on biological processes and molecular functions ($p < 0.05$, Log2 Fold change > 1). The size of the dots indicates the number of proteins in each category. The color of the dots corresponds to the p-adjust values. The x-axis represents the gene ratio. Right: Top enriched GO terms for downregulated proteins in each comparison based on biological processes and molecular functions ($p < 0.05$, Log2 Fold change < -1). The size of the dots indicates the number of proteins in each category. The color of the dots corresponds to the p-adjust values. The x-axis represents the gene ratio.

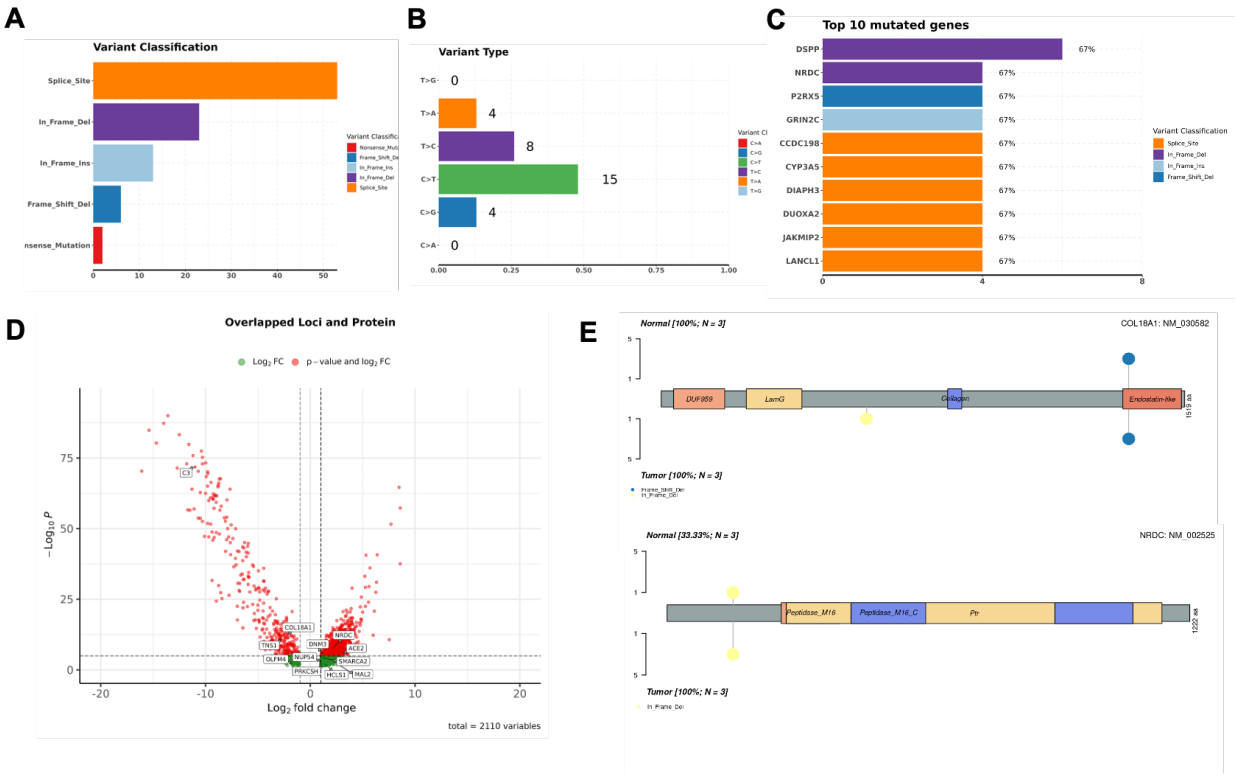

Figure 4. WES analysis of six paired invasive and adjacent normal TNAC samples (A) Overview of SNP site types: The majority of SNP sites are splice sites. (B) Overview of SNV type. Most SNVs were C>T. (C) Top ten mutated genes. The bar graph also shows the number of mutations (x-axis) and the percentage of patients with the mutations. (D) Overlapped loci with proteomics result. Twelve genes are mutated while proteins are significantly changed ($p < 0.01$, $|\text{Log2 Fold Change}| > 1$) in the proteomic dataset. (E) Schematic representation of point mutations in COL18A1 and NRDC in three TNAC patients (Tumor vs Normal). The lollipop indicates the mutation sites, with the color representing the classification of SNP and the length of the stick indicating the number of patients with the mutations.

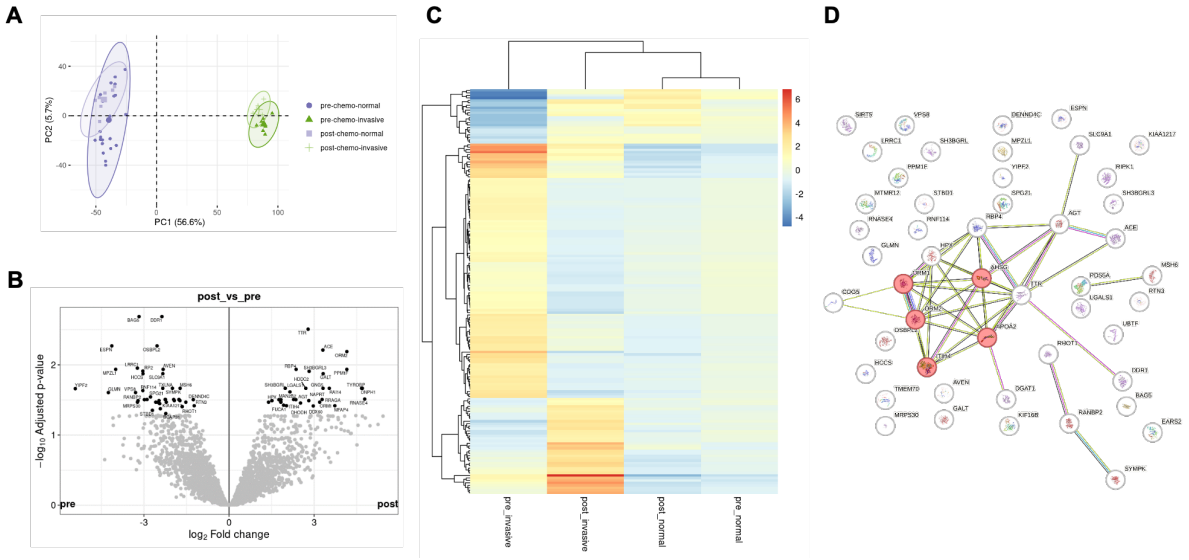

Figure 5: Proteomics landscape of chemotherapy in-sensitive TNAC patients. (A) PCA visualization of normal tissues from 24 patients not receiving chemotherapy (pre-chemo normal), and normal/tumor tissues from 7 patients before or after neoadjuvant chemotherapy (post-chemo normal, pre-chemo invasive, and post-chemo invasive). Every dot on the graph symbolizes a sample, and the percentage value shows the amount of variance explained. (B) Volcano plot of differentially expressed proteins in invasive tissues after therapy. The plots depict log2 Fold change versus –log10 p-value. The dark black dots are proteins defined as significantly changed ($p < 0.05$, |Log2 Fold change| > 1). (C) Heatmap of differentially expressed proteins shows slight differences after chemotherapy for invasive tumor tissues. Each cell illustrates the standardized score for protein abundance in a particular sample. The color scheme corresponds to the standardized score reflecting protein expression. (D) Protein-protein interaction network of up-regulated proteins after chemotherapy. Each circle represents a protein. Lines represent protein-protein interaction. Red nodes represent proteins involved in acute inflammatory response (GO:0002526).

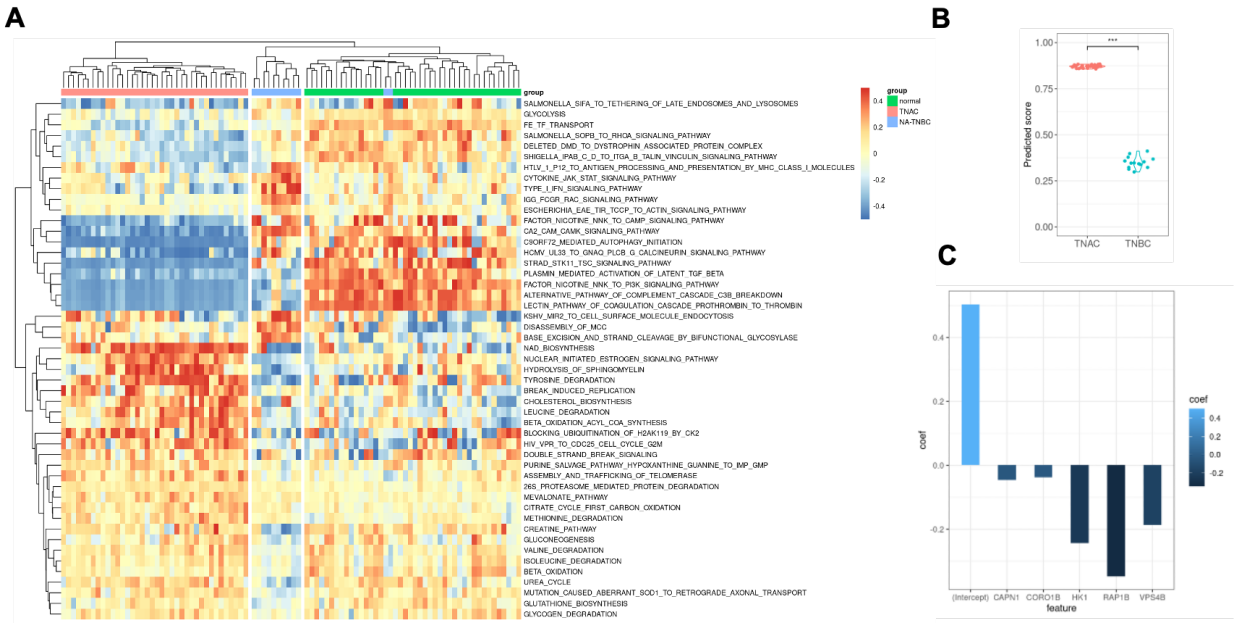

Figure 6: TNAC and NA-TNBC proteomics differentiation. (A) GSVA depicting the enrichment of different KEGG pathways in TNAC and NA-TNBC invasive tissues. The heatmap color indicates the degree of differentially expressed pathways. (B) Predicted scores of TNAC and (NA)-TNBC samples. A higher score represents a higher probability that the sample belongs to TNAC. A perfect classifier can be created based on the predicted score. (C) Coefficients of the model of a core protein panel (CAPN1, CORO1B, HK1, RAP1B, VPS4B) to differentiate TNAC and NA-TNBC.