A 15-Gene Predictive Signature for Pathologic Complete Response in Triple-Negative Breast Cancer: Validation Across Institutional and TCGA Cohorts

Ming-Feng Hou[1], Mei-Ren Pan[2], Sin-Hui Moi[2], Fang-Ming Chen[1], Chi-Wen Luo[1]

[1]Division of Breast Oncology and Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan,
[2]Graduate Institute of Clinical Medicine, Kaohsiung Medical University, Kaohsiung city, Taiwan.

SAN ANTONIO BREAST CANCER SYMPOSIUM®

PS4-04-01

Purpose

Pathologic complete response (pCR) to neoadjuvant chemotherapy (NAC) is a strong prognostic marker in triple-negative breast cancer (TNBC). However, predictive biomarkers for pCR remain limited. This study aimed to identify gene expression profiles associated with pCR in TNBC to aid treatment stratification.

Materials and Methods

Tumor samples were obtained from 16 TNBC patients treated with NAC at Kaohsiung Medical University Hospital (KMUH), including 5 pCR and 11 non-pCR cases. RNA-seq was performed, and differentially expressed genes (DEGs) were identified using DESeq2. Expression profiles were compared with a validation cohort of 27 chemotherapy-responsive TNBC cases from TCGA. Overlapping DEGs were identified via Venn diagram analysis. Drug-gene interaction databases were queried to assess therapeutic relevance.

Results

A total of 175 DEGs were found in the KMUH cohort, with 146 upregulated and 29 downregulated in non-pCR tumors. Fifteen DEGs showed consistent expression patterns between KMUH non-pCR tumors and both KMUH pCR and TCGA responder cohorts. These genes were enriched in pCR samples and may serve as predictive biomarkers. Several candidates were identified as potentially druggable, suggesting relevance for therapeutic targeting.

Conclusions

We identified a 15-gene signature associated with pCR in TNBC, validated across independent cohorts. These biomarkers may inform treatment decisions, improve patient stratification, and offer new directions for targeted therapy development in chemoresistant TNBC.

Schematic workflow for the identification of candidate genes associated with chemotherapy response in TNBC patients

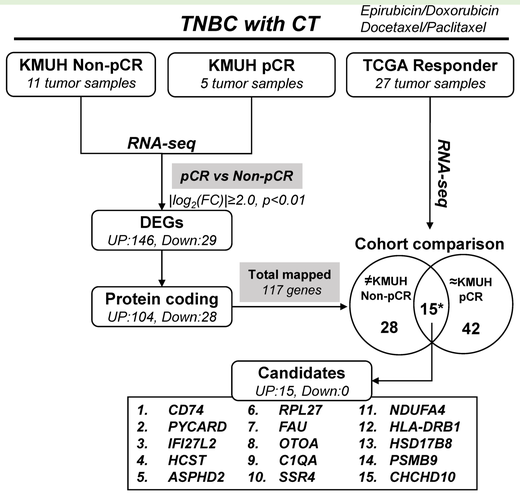


Figure 1. RNA-seq analysis was performed on the KMUH cohort to identify differentially expressed protein-coding genes. Cross-comparison with the TCGA responder cohort (n=27) narrowed the selection to 15 candidate genes potentially associated with chemotherapy resistance.

Table 1. Baseline characteristics of KMUH and TCGA cohort

Characteristics	KMUH				TCGA
	Overall	Non-pCR	pCR	*P*	Responder
Cases no.	16	11	5		27
Diagnosis age	58.6 ± 9.2	55.5 ± 9.3	65.4 ± 4.0	0.061	52.9 ± 10.0
Tumor stage				1.000	-
1	1 (6.3%)	1 (9.1%)	0 (0%)		0 (0.0%)
2	7 (44%)	4 (36%)	3 (60%)		4 (14.8%)
3	7 (44%)	5 (45%)	2 (40%)		22 (81.5%)
4	1 (6.3%)	1 (9.1%)	0 (0%)		1 (3.7%)
Tumor size	1.9 ± 2.3	2.1 ± 2.4	-	0.144	-
Margin				0.083	
Free	14 (88%)	11 (100%)	3 (60%)		-
Unknown	2 (13%)	0 (0%)	2 (40%)		
Grade				0.036	-
2	6 (38%)	2 (18%)	4 (80%)		
3	10 (63%)	9 (82%)	1 (20%)		
Ki67				0.313	-
<20	1 (6.3%)	0 (0%)	1 (20%)		
≥20	15 (94%)	11 (100%)	4 (80%)		
Treatment					
Surgery	15 (94%)	11 (100%)	4 (80%)	0.313	27 (100%)
Radiotherapy	9 (56%)	6 (55%)	3 (60%)	1.000	18 (66.7%)
Target therapy	7 (44%)	6 (55%)	1 (20%)	0.308	0 (0.0%)
Hormone therapy	4 (25%)	4 (36%)	0 (0%)	0.245	3 (11.1%)
Survival outcome					
Progressed	1 (6.3%)	1 (9.1%)	0 (0%)	1.000	0 (0.0%)
Died	3 (19%)	3 (27%)	0 (0%)	0.509	1 (3.7%)

P-value estimated using Wilcoxon rank sum test and Fisher's exact test.

Transcriptomic profiling and identification of differentially expressed genes associated with chemotherapy response in the KMUH cohort

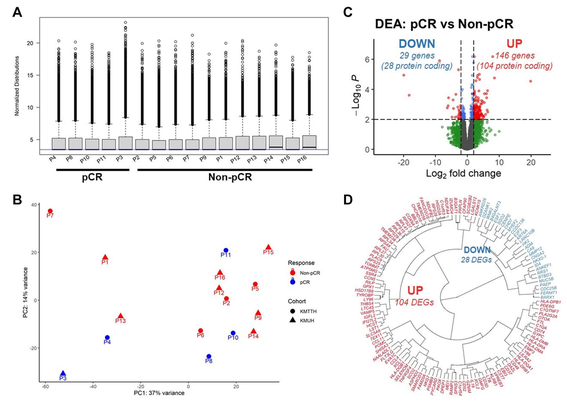


Figure 2. RNA-seq analysis of the KMUH cohort. (A) Normalized mRNA expression distributions across all samples. (B) PCA plot. Shapes denote sample batches (circle: batch 1; triangle: batch 2) and colors indicate clinical response (Blue: pCR; Red: Non-pCR). (C) Volcano plot illustrating DEGs comparing Non-pCR vs. pCR groups. (D) Circular dendrogram clustering the identified protein-coding DEGs based on expression profiles.

Table 2. Summary of 15 candidate DEGs for pCR estimation.

Gene	Regulation[1v2]	KMUH (Non-pCR)[1]	KMUH (pCR)[2]	$\log_2$(FC)[1v2]	p-value[1v2]	q-value[1v2]	P^{2v1}	TCGA (Responder)[3]	P^{3v1}	P^{3v2}
CD74	UP	3.18 (2.96, 3.27)	3.51 (3.32, 3.83)	2.03	<0.001	0.003	0.019	3.44 (3.11, 3.57)	0.030	0.263
PYCARD	UP	1.00 (0.81, 1.07)	1.40 (1.37, 1.51)	2.29	<0.001	0.005	0.002	1.53 (1.30, 1.82)	<0.001	0.418
IFI27L2	UP	0.89 (0.75, 0.99)	1.46 (1.29, 1.48)	3.14	<0.001	<0.001	0.013	1.10 (0.93, 1.27)	0.016	0.068
HCST	UP	0.41 (0.31, 0.60)	1.03 (1.00, 1.21)	3.43	<0.001	<0.001	0.002	0.98 (0.74, 1.13)	<0.001	0.263
ASPHD2	UP	0.18 (0.08, 0.29)	0.52 (0.34, 0.58)	2.05	<0.001	0.004	0.027	0.65 (0.53, 0.82)	<0.001	0.263
RPL27	UP	2.45 (2.33, 2.55)	2.77 (2.62, 2.79)	2.91	<0.001	0.007	0.038	2.59 (2.49, 2.82)	0.013	0.545
FAU	UP	2.21 (2.05, 2.34)	2.56 (2.42, 2.61)	3.16	<0.001	0.005	0.009	2.45 (2.39, 2.61)	<0.001	0.579
OTOA	UP	-0.43 (-0.88, -0.34)	0.30 (0.03, 0.33)	4.26	<0.001	<0.001	0.038	-0.26 (-0.41, -0.05)	0.005	0.122
C1QA	UP	1.86 (1.76, 2.14)	2.46 (2.33, 2.57)	2.29	<0.001	0.005	0.002	2.22 (2.04, 2.37)	0.013	0.053
SSR4	UP	2.23 (2.04, 2.27)	2.48 (2.45, 2.63)	2.98	<0.001	0.006	0.013	2.35 (2.16, 2.45)	0.049	0.136
NDUFA4	UP	2.02 (1.84, 2.09)	2.13 (2.08, 2.14)	3.26	<0.001	0.006	0.038	2.20 (2.13, 2.30)	<0.001	0.448
HLA-DRB1	UP	1.83 (1.53, 2.11)	2.29 (2.23, 2.50)	2.67	<0.001	0.003	0.019	2.52 (2.29, 2.79)	<0.001	0.389
HSD17B8	UP	0.57 (0.45, 0.79)	0.98 (0.97, 1.04)	2.24	<0.001	0.008	0.009	0.81 (0.60, 1.00)	0.023	0.241
PSMB9	UP	1.45 (1.27, 1.69)	1.91 (1.88, 2.13)	2.37	<0.001	0.006	0.005	1.73 (1.52, 2.03)	0.045	0.263
CHCHD10	UP	1.18 (0.98, 1.24)	1.33 (1.30, 1.42)	3.01	<0.001	0.006	0.027	1.38 (1.18, 1.65)	0.019	0.763

[1,2,3] Expression of each genes were transformed into $\log_{10}$(CPM).
[1v2] Differential expression analysis was performed using expression between KMUH (Non-pCR)[1] and KMUH (pCR)[2].
[2v1] Pairwise statistic results between KMUH (pCR)[2] and KMUH (Non-pCR)[1] based on Wilcoxon rank-sum test.
[3v1] Pairwise statistic results between TCGA (Responder)[3] and KMUH (Non-pCR)[1] based on Wilcoxon rank-sum test.
[3v2] Pairwise statistic results between TCGA (Responder)[3] and KMUH (pCR)[2] based on Wilcoxon rank-sum test

A 15-Gene Signature Predicts pCR and Reveals Druggable Targets

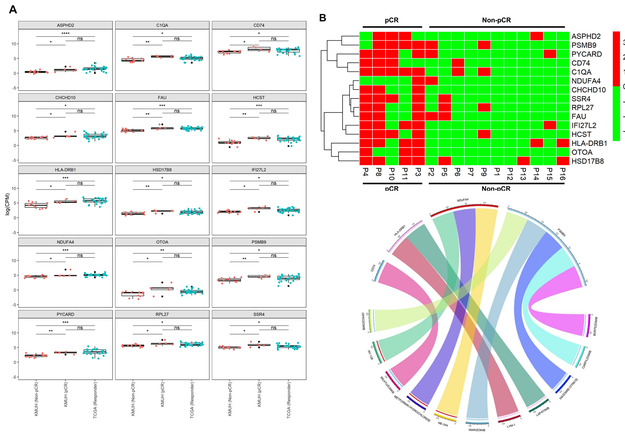


Figure 3. A 15-Gene Signature and Its Druggable Targets (A) Expression of the 15 pCR-associated genes across KMUH pCR, KMUH non-pCR, and TCGA responder groups. (B) Heatmap showing consistent up-regulation in pCR and down-regulation in non-pCR samples. (C) Chord diagram of drug–gene interactions for four druggable DEGs.

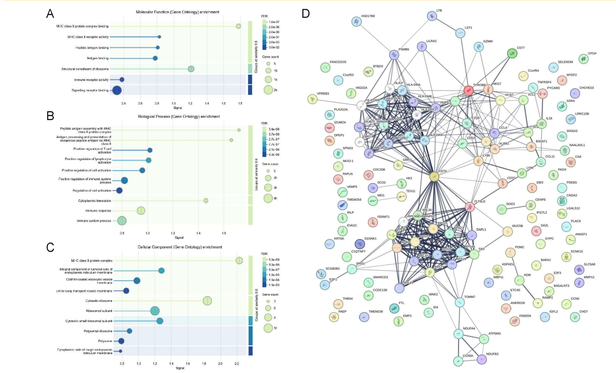


Figure 4. The GO enrichment and PPI analysis results for all DEGs. Top 10 enriched results of (A) molecular function, (B) biological process, and (C) cellular component for all DEGs. (D) The PPI network of all DEGs.