Supplementary Data

Supplementary Methods

Tissue Processing - The tumor tissue underwent macro-dissection to enhance the tumor content of the study material. A frozen aliquot of tissue at dry ice temperature (-78^{0}C) was placed in the center of a frozen section cryo-mold and then surrounded by room temperature OCT matrix and immediately frozen on a cryostat chuck (-20^{0}C). The tissue specimen was not permitted to thaw except at the superficial edges at the OCT-tissue interface. After the chuck with the specimen was mounted in the cryostat, frozen sections were removed in order to efface the block and demonstrate the whole aliquot. A frozen section was cut at 8μ stained with hematoxylin and eosin (H&E) and a coverslip applied. A Sharpie black pen was used to orientate the aliquot by marking the OCT adjacent to the specimen and the coverslip over the frozen section was similarly marked with the Sharpie pen so that the orientation of the H&E frozen section could be matched closely to the OCT block. The pathologist (WEG) then reviewed the orientated H&E frozen section and marked on the coverslip using a blue Sharpie pen (ultra-fine point) the areas of the specimen that will be removed (e.g., areas of uninvolved breast and/or of leukocytic infiltration) to enrich the specimen with malignant cells. The % tumor and % malignant cells are visually estimated and recorded. Using the marked H&E slide as a guide, the pathologist, with a single edge razor blade, cut away unnecessary and unwanted tissue from the orientated OCT specimen. The remaining specimen was removed from the OCT block and if too thick, was trimmed to a thickness of less than 1 mm and OCT was removed from the edges of the specimen. The trimmed specimen was then wrapped in aluminum foil,

placed into a tissue cassette and returned to a -80^{0}C freezer until shipped on dry ice to HudsonAlpha Institute for Biotechnology. The de-identified shipped tumor specimens had >50% tumor nuclei.

RNA-seq - Upon arrival, the 47 frozen breast tumor tissue specimens were weighed and transferred to 15 mL conical tubes containing 100uL of ceramic beads (Lysing Matrix D from MP Biomedicals). Lysis buffer composed of RLT Buffer (Qiagen) supplemented with 1% BME was added so that each tube containing 35 uL of buffer for each milligram of tissue. To homogenize the tissue the conical tubes containing tissue, ceramic beads, and buffer were then shaken in a MP Biomedicals FastPrep machine for 90 seconds at 6.5 meters per second. The homogenized tissue was stored at -80°C. For analysis, total RNA was extracted from 350 uL of tissue homogenate (equivalent to 10 mg of tissue) using the Norgen Animal Tissue RNA Purification Kit with the optional Proteinase K treatment and on-column DNAse treatment according to the manufacturer's instructions (Norgen Biotek Corporation). Total RNA was quantified using the Qubit RNA Assay Kit and the Qubit 2.0 fluorometer (Invitrogen). RNA-seq libraries for each sample were constructed from 250 ng total RNA using the polyA selection and transposase-based non-stranded library construction (Tn-RNA-seq) described previously (1).

RNA-seq data analysis - RNA-seq read pairs (50 million per sample) were aligned to the transcriptome using TopHat v1.4.1 (2) and GENCODE version 9 (3) was used as the transcript reference. Bedtools (4) was used to calculate the read count per reference gene. Gene expression values (Fragments Per Kilobase of transcript Per Million reads, FPKMs) were

calculated for each GENCODE transcript using Cufflinks 1.3.0 with the –u option (5). We performed unsupervised clustering on normalized gene read counts to identify subclusters of samples within our dataset using the R ConsensusClusterPlusR package (6)(http://www.bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html). The following k-means clustering options were used: maxK=10, reps=1000, clusterAlg="km", distance="Euclidean". Analysis of the Consensus Cumulative Distribution Function (CDF) and Delta Area Plot produced by the software indicated that at three clusters (k=3) the CDF reaches a maximum and there is diminishing increases in consensus at higher values of k.

The SAMseq function, part of the same R library (7) (http://cran.fhcrc.org/web/packages/samr/index.html), was used to perform supervised analysis to identify genes that are significantly differentially expressed between consensus Clusters 1, 2, and 3. Each tumor was classified as belonging to Cluster 1, Cluster 2, or Cluster 3 and the gene read counts were analyzed using the following SAMseq options: resp.type="Multiclass", nperms=1000, fdr.output=0.05. The genes that were significantly differentially expressed between classes (q-value<5%), with a positive contrast in Cluster 1 and negative contrast in the other two clusters were identified as specifically highly expressed in Cluster 1. This process was repeated for Cluster 2 and Cluster 3.

SAMseq (7)was also used to identify genes that are significantly differentially expressed between tumors from patients who relapsed and tumors from patients who did not relapse. Gene read count values were the input for SAMseq and the following options were used:

resp.type="Two class unpaired", fdr.output=0.05, nperms=1000. Genes with q-values <5% were considered significant.

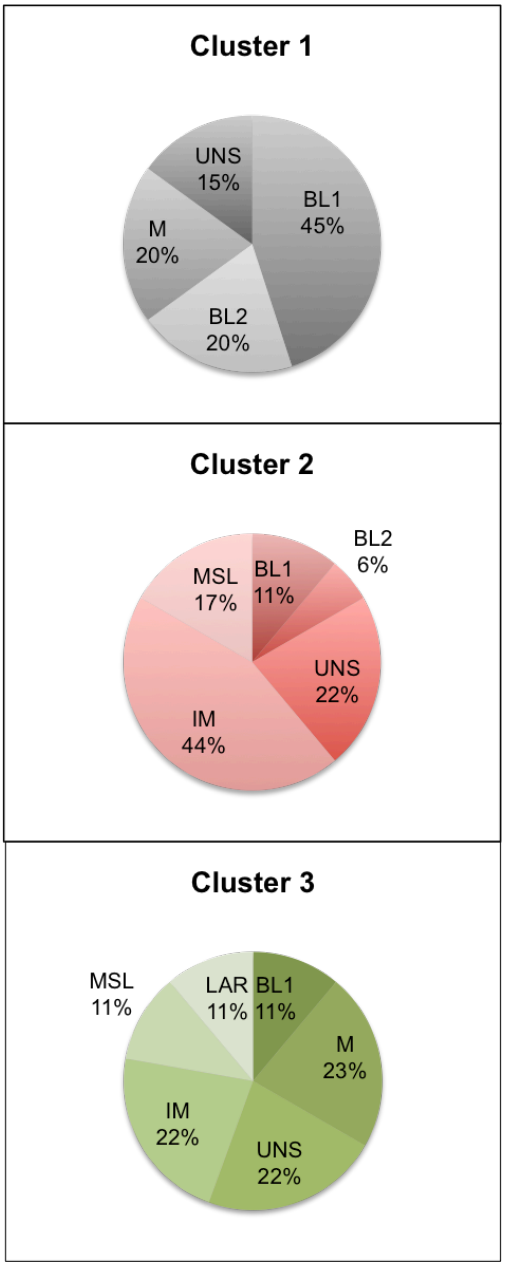
A heatmap depicting the expression of differentially expressed genes across patient samples was created using the Heatplus library in R (http://www.bioconductor.org/packages/release/bioc/html/Heatplus.html). The heatmap gene expression values were FPKMs with a plus 1 pseudocount added for accurate log base 2 transformation, i.e. log2 (FPKMs+1). Euclidean distance and complete linkage were used to cluster the genes. Patient samples were ordered based upon whether they experienced a relapse or not.

Kaplan-Meier curves and survival analysis was performed using RNA-seq FPKM values and an R script (http://kmplot.com/analysis/studies/Supplemental%20R%20script%201.R) (8). Patients were split into two groups using the optimal threshold, which split the lower tertile from the upper two tertiles.

Immunohistochemistry - Five archived de-identified TNBC tumor cases with sufficient tumor embedded in FFPE for immunohistochemical analysis were selected. Five 4 microns thick sections were cut from each FFPE block and placed on positively charged slides (plus slides). The slides were melted at 60°C for 30 minutes and placed in a Ventana automated stainer (Ventana Medical Systems, Inc.) for de-paraffinization with EZ Prep, treatment with Cell Conditioning Solution 1 (pH 8.0) 95°C for 64 minutes, and primary antibody incubation at a 1:320 dilution for 40 minutes at 37°C. The primary antibody for detection of CD74 was

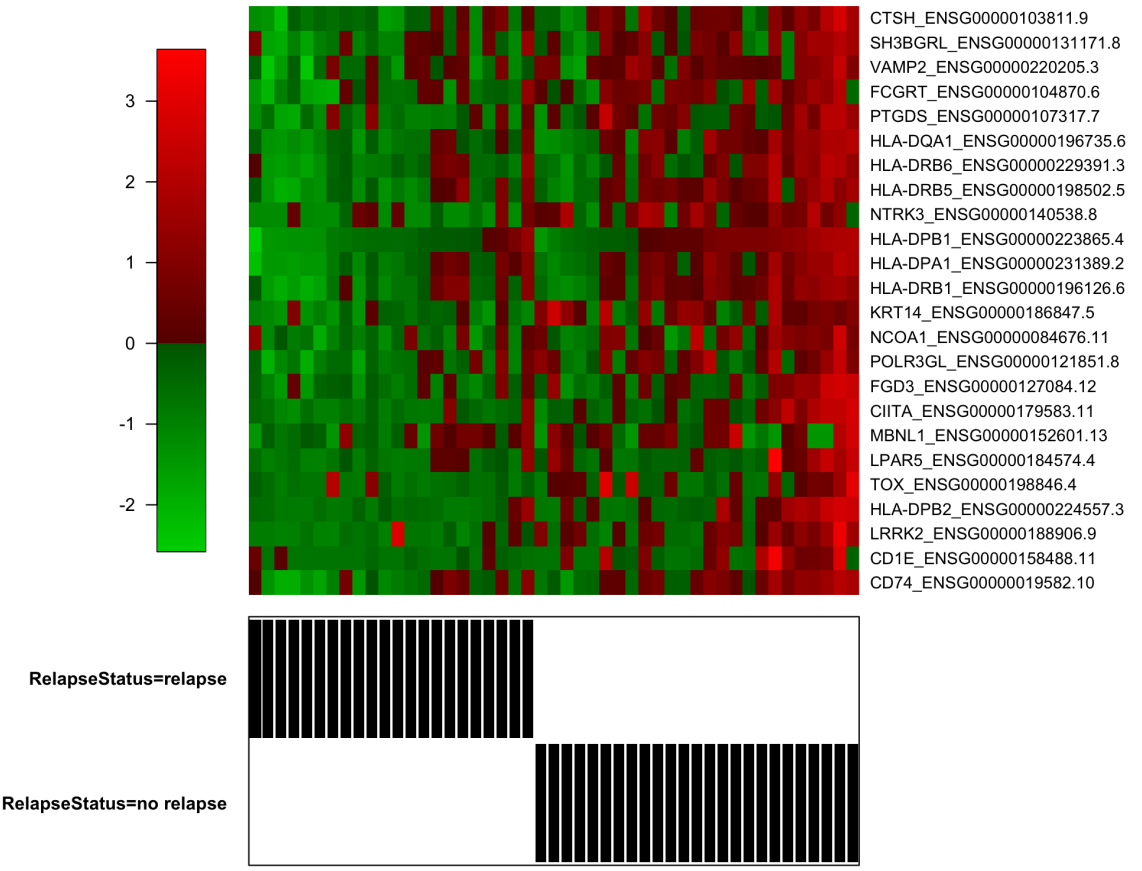
obtained from Leica/Novocastra (catalog number NCL-LN2) and the primary antibody for detection of HLA-DPB1 was obtained from Sigma-Aldrich (catalog number HPA011078). The ultraView Universal DAB detection kit (Ventana Medical Systems, Inc.) was used to detect primary antibody staining followed by a hematoxylin counterstain for 8 minutes. The slides were washed with a mixture of water and Dawn soap to remove oil from the automated instrument, followed by water washes to remove the soap. A 30 second iodine wash was performed to remove any metal precipitates from the fixation, followed by a 30 second wash in sodium thiosulfate to remove residual iodine. The slides were then dehydrated in graded alcohols (70%, 95%, and 100%) followed by 4 xylene washes. Coverslips were applied and the slides were air-dried. An anatomic pathologist reviewed the stained slides and estimated the fraction of positive tumor cells and described the localization of the staining.

Supplementary Figure 1



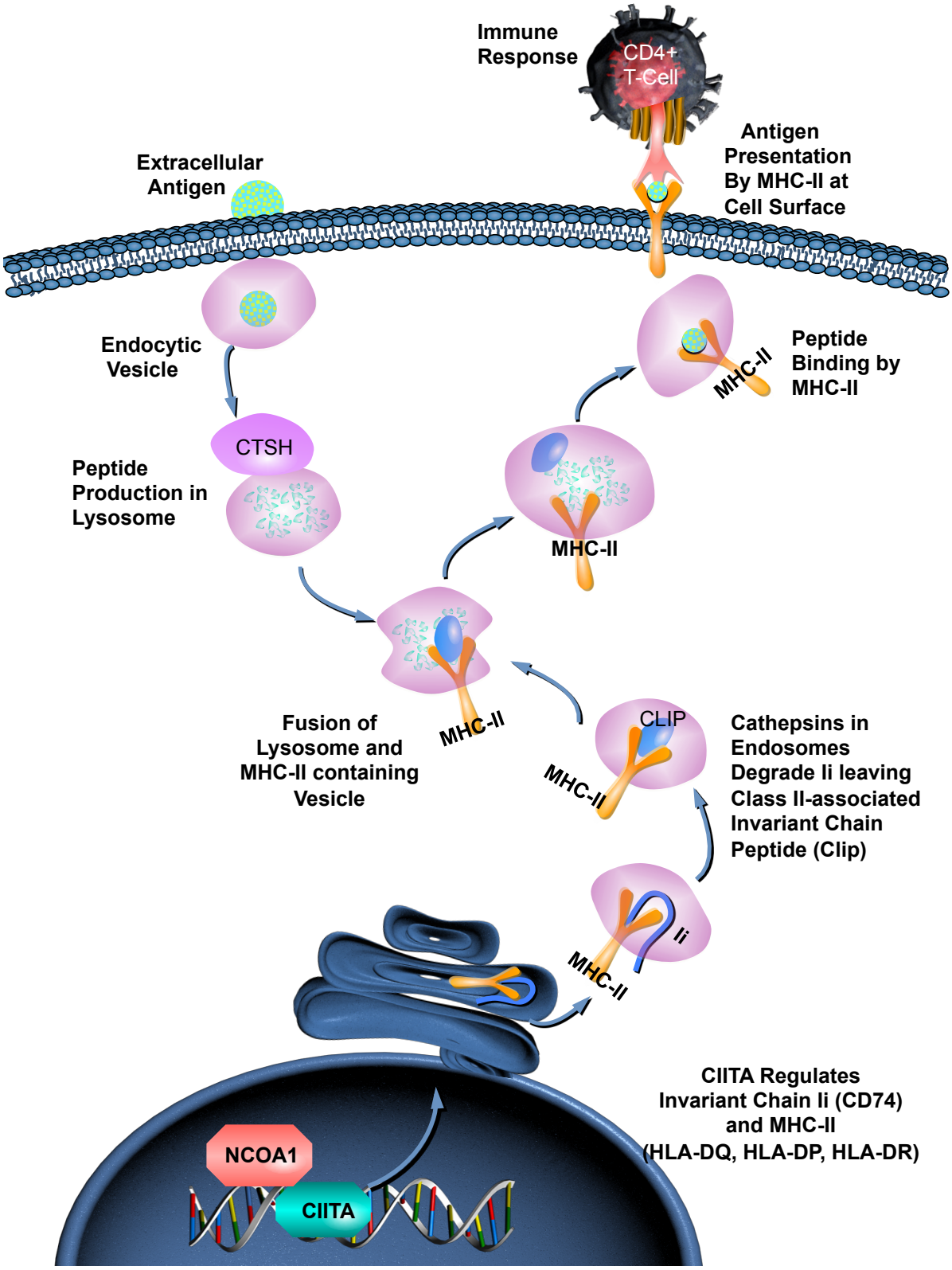
Supplementary Figure 1. Overlap between clusters identified in our data and TNBC subtypes described in previous studies. Each of the clusters we identified contained multiple TNBC subtypes and the TNBC subtypes were represented in multiple clusters. TNBCType Subtypes are: basal-like (BL1 and BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem–like (MSL), luminal androgen receptor (LAR), and unassigned (UNS).

Supplementary Figure 2



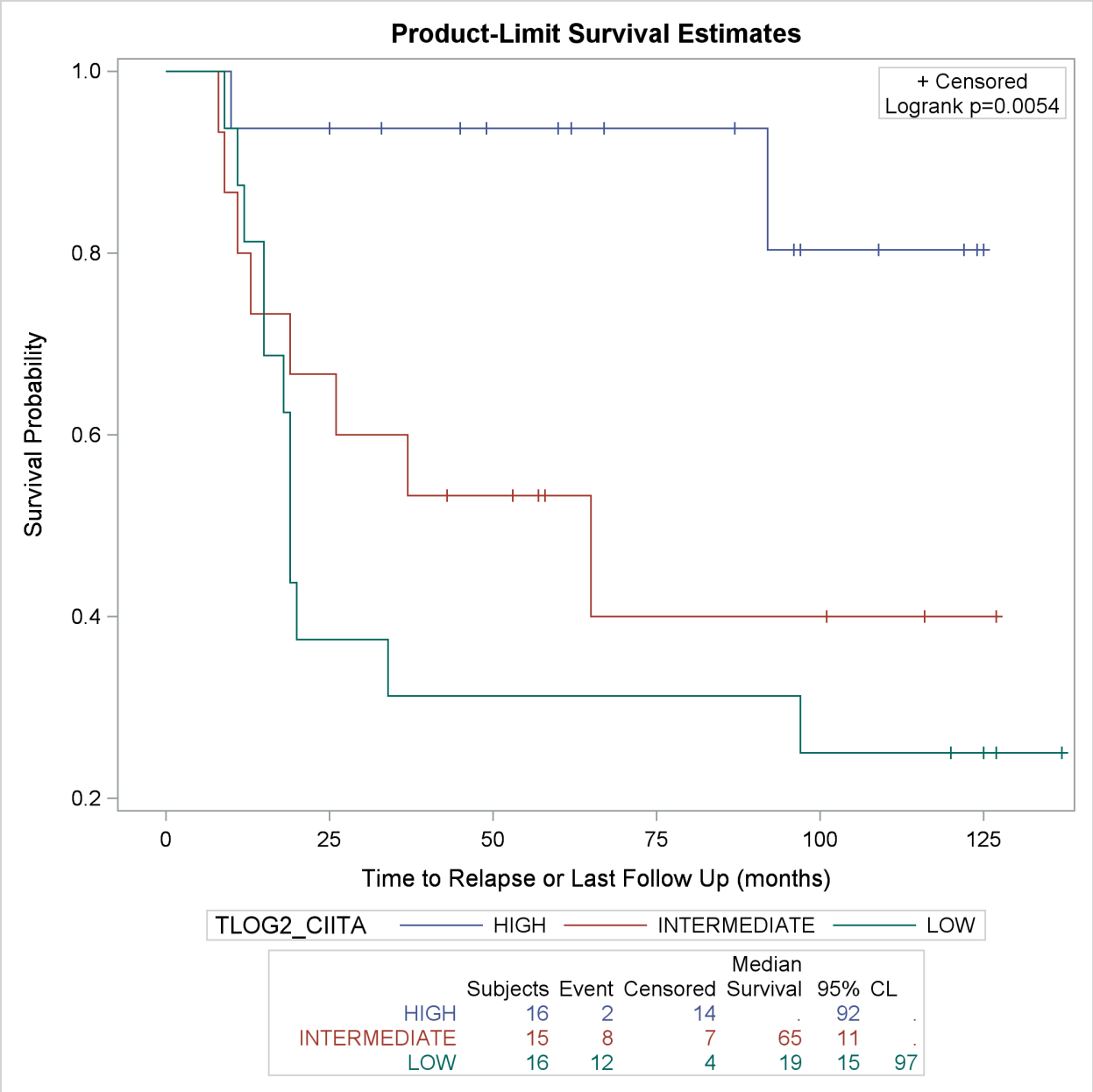
Supplementary Figure 2. Heatmap of the normalized gene expression values of each of the 24 prognostic genes in each of the 47 patient's tumors. Red indicates higher expression and green indicates lower expression. Patients who relapsed are grouped on the left, and patients who did not relapse are grouped on the right. Gene identifiers on the right include the common gene symbol followed by the unique gene identifier used by Ensembl.

Supplementary Figure 3



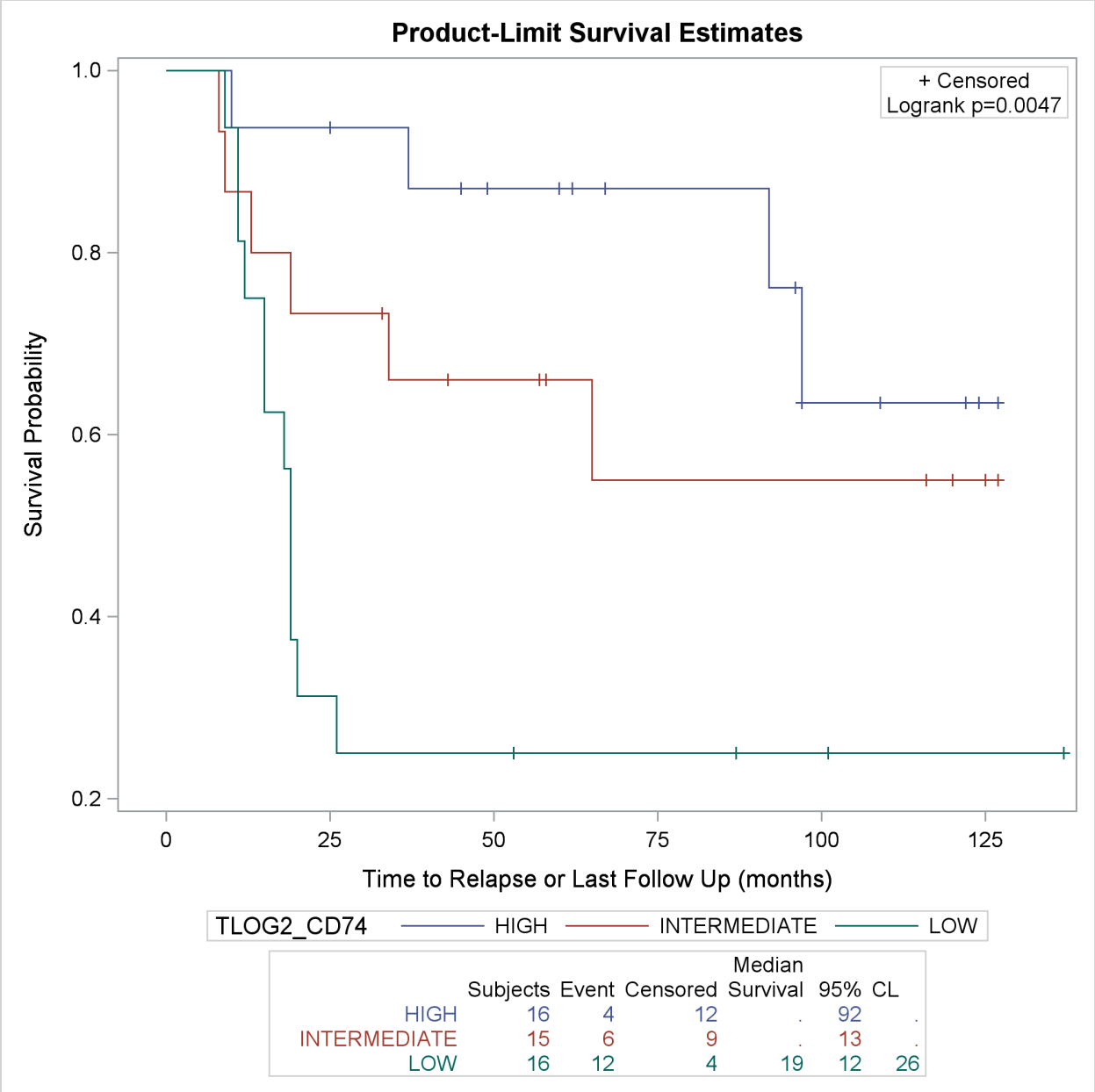
Supplementary Figure 3. Illustration of TNBC prognostic genes in the MHC II pathway

Supplementary Figure 4A



	Subjects	Event	Censored	Median Survival	95% CL	
HIGH	16	2	14	.	92	.
INTERMEDIATE	15	8	7	65	11	.
LOW	16	12	4	19	15	97

Supplementary Figure 4B



Supplementary Figure 4. Kaplan-Meier PFS curves of patients with expression levels (tertiles) of CIITA and CD74. (A) High (blue solid) intermediate (red), and low (blue interrupted) levels of CIITA; log-rank p = 0.0054. (B) High (blue solid line), intermediate (red), and low (blue interrupted) levels of CD74; log-rank p = 0.0047.

Supplementary Tables

Supplementary Table 1. Cluster Content and Outcomes

	Relapse	Basal-like Subtype	Number of Genes that are Specifically Upregulated (FDR <0.01)	Significantly Enriched Gene Families and Pathways	Cluster Descriptor
Cluster 1	12/20 (60%)	18/20 (90%)	806	Mitotic G1/S transition cell cycle checkpoints actin and tubulin folding	Basal-like
Cluster 2	3/17 (18%)	13/17 (76%)	839	Immunoregulatory interactions T cell receptor signaling complement cascade	Immuno-modulatory
Cluster 3	7/10 (70%)	7/10 (70%)	71	Genes in cytogenetic band chr16p13	Unclassified

Supplementary Table 2. Correlation Coefficient and Hazard Ratio of MHCII Genes										
				HLA						
	CIITA	**CD74**	**CTSH**	**DPA1**	**DPB1**	**DPB2**	**DQA1**	**DRB1**	**DRB5**	**DRB6**
CIITA	1	0.84	0.68	0.84	0.92	0.86	0.83	0.71	0.71	0.71
CD74	0.83	1	0.86	0.98	0.86	0.79	0.92	0.92	0.87	0.87
CTSH	0.68	0.86	1	0.85	0.75	0.72	0.82	0.74	0.74	0.74
DPA1	0.84	0.98	0.85	1	0.88	0.78	0.93	0.89	0.82	0.82
DPB1	0.92	0.86	0.75	0.86	1	0.92	0.86	0.75	0.78	0.78
DPB2	0.86	0.77	0.72	0.78	0.92	1	0.76	0.65	0.70	0.70
DQA1	0.83	0.92	0.82	0.93	0.86	0.76	1	0.88	0.84	0.84
DRB1	0.71	0.92	0.74	0.89	0.75	0.65	0.88	1	0.74	0.94
DRB5	0.71	0.87	0.74	0.82	0.78	0.70	0.84	0.94	1	0.85
DRB6	0.71	0.87	0.74	0.82	0.78	0.70	0.84	0.94	0.85	1
	Progression Free Survival Hazard Ratio (> vs < median value)									
HR* (95% CI)	0.167 (0.056-0.496)	0.349 (0.141-0.865)	0.225 (0.083-0.611)	0.292 (0.114-0.751)	0.234 (0.086-0.638)	0.241 (0.088-0.654)	0.280 (0.109-0.719)	0.202 (0.074-0.552)	0.190 (0.069-0.523)	0.263 (0.101-0.684)
p Value **	0.0013	0.0230	0.0035	0.0106	0.0045	0.0052	0.0081	0.0018	0.0013	0.0062

*** Above vs below median**
**** Univariate Cox Regression Model**

Supplementary Data References:

1. Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, et al. Transposase mediated construction of RNA-seq libraries. Genome research. 2012 Jan;22(1):134-41.
2. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009 May 1;25(9):1105-11.
3. Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, et al. GENCODE: producing a reference annotation for ENCODE. Genome Biol. 2006;7 Suppl 1:S4 1-9.
4. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2.
5. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010 May;28(5):511-5.
6. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010 Jun 15;26(12):1572-3.
7. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Statistical methods in medical research. 2013 Oct;22(5):519-36.
8. Mihaly Z, Kormos M, Lanczky A, Dank M, Budczies J, Szasz MA, et al. A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. Breast cancer research and treatment. 2013 Jul;140(2):219-32.