Supplemental Results

K-means Clustering of a subset of five datasets that were identified as TNBC by IHC. Since TNBCs are not identified by GE in the clinical setting, we performed a similar GE analysis on a dataset from TNBCs identified using IHC (n=183) complied from five datasets: GSE7904 (n= 22), GSE19615 (n= 34), GSE20194 (n= 67), E-TABM-158 (n= 30) and GSE22513, GSE-XXX (n= 30). This compilation included 136 tumors (74.3%) that were identified by our bimodal filtering analysis (Table S4). K-means clustering performed on the most differentially expressed genes (SD> 0.8) resulted in the identification of 5 TNBC subtypes (Figure S4). One subtype contained only three tumors and analysis of GE showed these tumors to have high levels of ER and PR (Figure S4B). These tumors were considered to be false negatives by IHC and were removed from further analysis. Comparison of these IHC-identified TNBC samples using genes differentially expressed from the six TNBC subtypes revealed similar patterns of gene enrichment (Figure S5).

Supplemental Methods

Immunostaining. Both formalin fixed paraffin embedded (FFPE) and frozen tissue were used for immunohistochemical studies. When frozen tissue was used for immunohistochemistry, staining was performed on sections taken from the same tissue block from which RNA was isolated for microarray. AR, Ki67, EGFR and CK 5/6 expression were evaluated in frozen tissue using the DAKO (Carpinteria, CA) antibodies: AR (clone AR411) at a 1:50 dilution, EGFR at 1:200, CK5/6 at 1:50 and Ki67 antibody (MIB1 clone) at a 1:200 dilution for 1 h at room temperature. FFPE tissue was subject to antigen retrieval with high pH buffer (pH 8.0) followed by overnight incubation with an AR (1:30) or Ki67 (1:75) antibody dilution overnight. AR expression was scored as both the percentage of tumor cells with nuclear staining as well as the intensity of staining (scored as 0-3+). An AR intensity score was calculated as follows: (AR intensity x 100) + % AR positive cells. For Ki67 the percentage of cells demonstrating nuclear staining at any intensity was recorded.

Immunoblotting. Cells were trypsinized, lysed and relative protein expression was determined by Western blot as previously described with the following antibodies; HSP90 monoclonal antibody, clone F-18 (Santa Cruz Biotechnology, Santa Cruz, CA) and the AR polyclonal antibody, SC-N20 (Santa Cruz Biotechnology).

Colony Formation. MFM-223, MDA-MB-453 and CAL-148 cells (3000 cells/well) were reverse-transfected with 1.25 pmole of siRNAs to AR [ON-TARGET plus SMARTpool, cat# L-003400-00 (Dharmacon, Lafayette, CO)] or non-targeting control (ON-TARGETplus Non-targeting Pool cat# D-001810-10-05) with 0.25 μL Dharmafect #3 (MFM-223), or 0.25 μL Dharmafect #1 (MDA-MB-453 and CAL-148). Colonies were stained and quantified 14 d following transfection using Cell Profiler 2.0 (Broad Institute, Cambridge, MA). Experiments were preformed in triplicate and error bars reflect standard deviation.

Figure S1

A

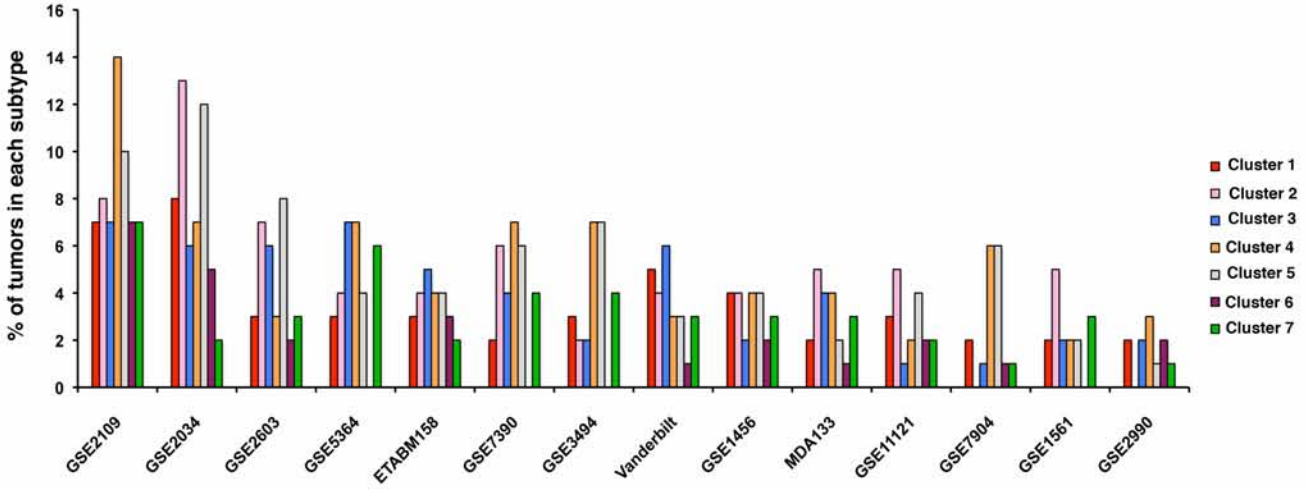


Figure S1. Random distribution of TNBC subtypes found within datasets. (A) The percent of tumors in each cluster is displayed across 14 studies that comprise the training dataset.

Figure S2

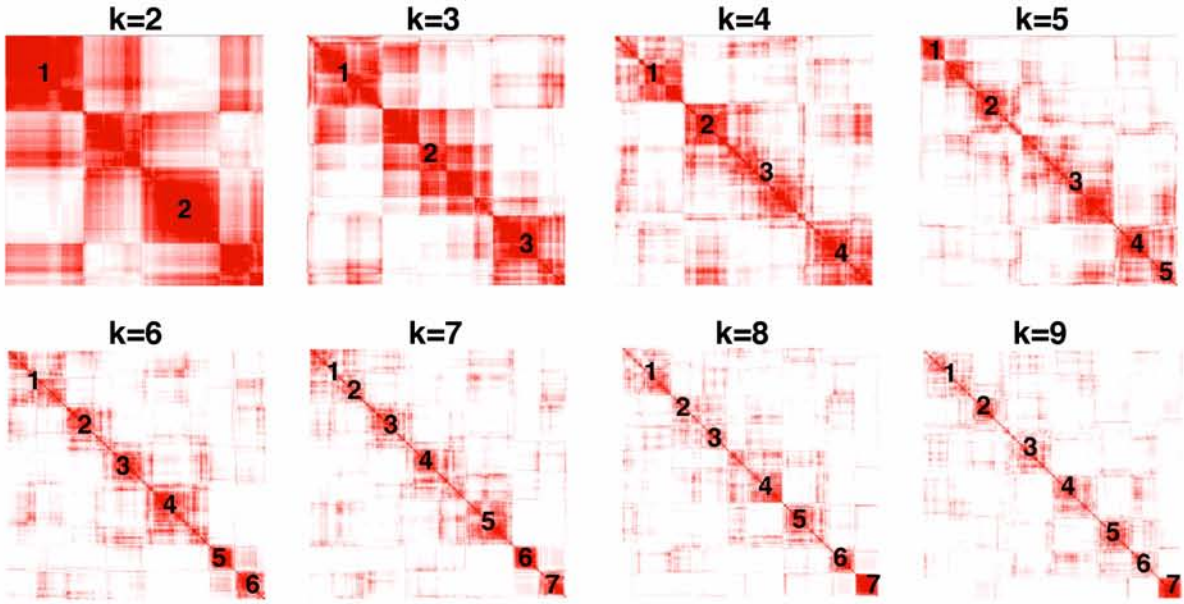


Figure S2. Consensus clustering of training and validation TNBC datasets (n= 587). Heat map displays consensus clustering results depicting the robustness of sample classification. Red areas indicate samples that frequently cluster with each other over multiple iterations (1000) of k-means clustering (k= 2 to k= 9).

Figure S3

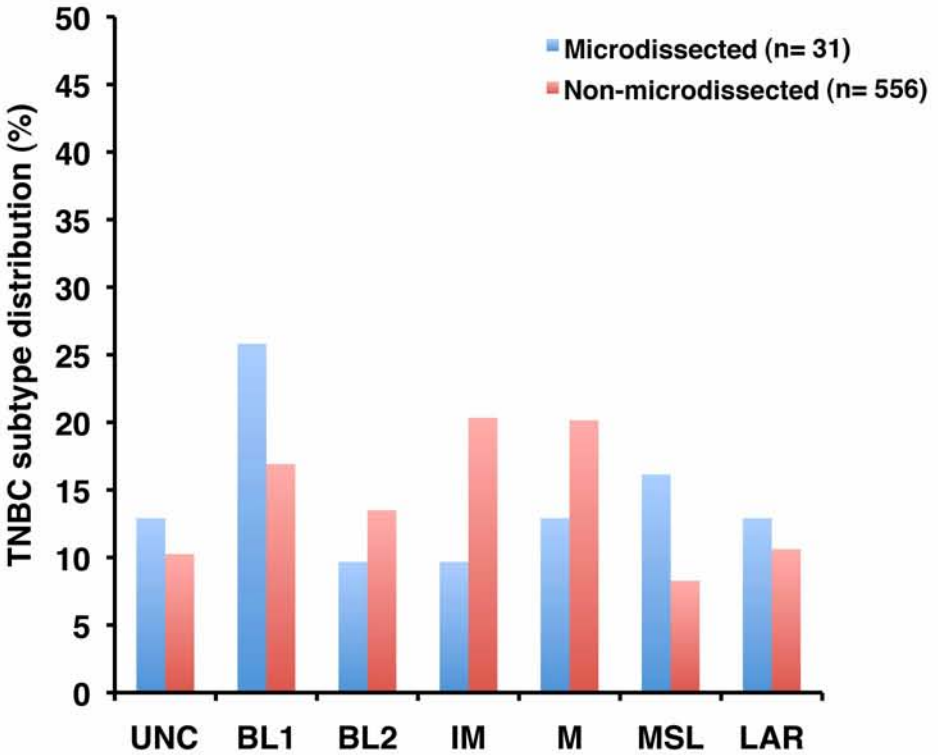
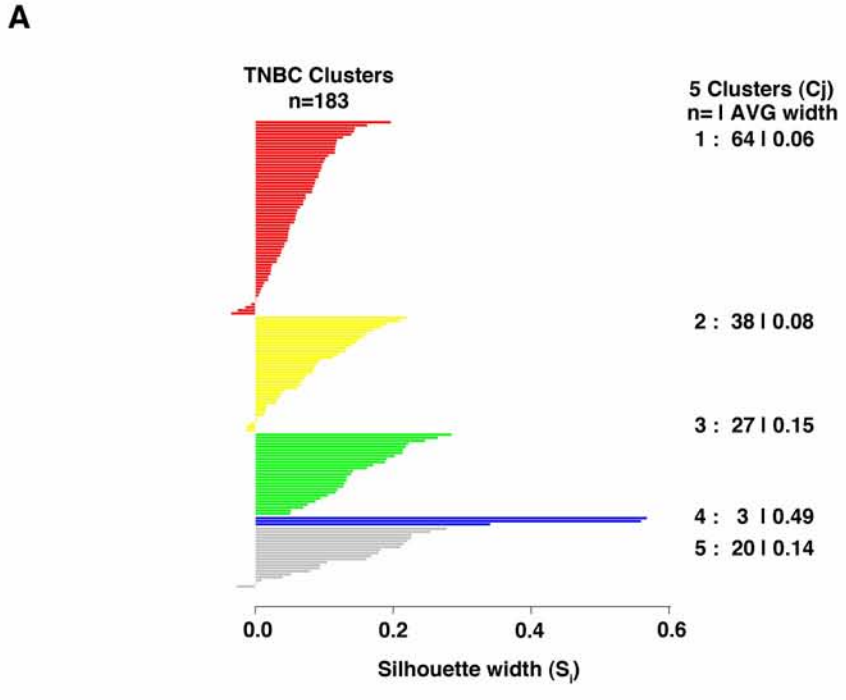


Figure S3. Gene expression profiles derived from tumors that were laser-capture microdissected represent all TNBC subtypes. Bar graph depicts the percentage of tumors that were microdissected (blue bars) present in the GSE584 (n=17) and Vanderbilt (n=14) datasets relative to non-microdissected tumors in the combined dataset (red bars).

Figure S4



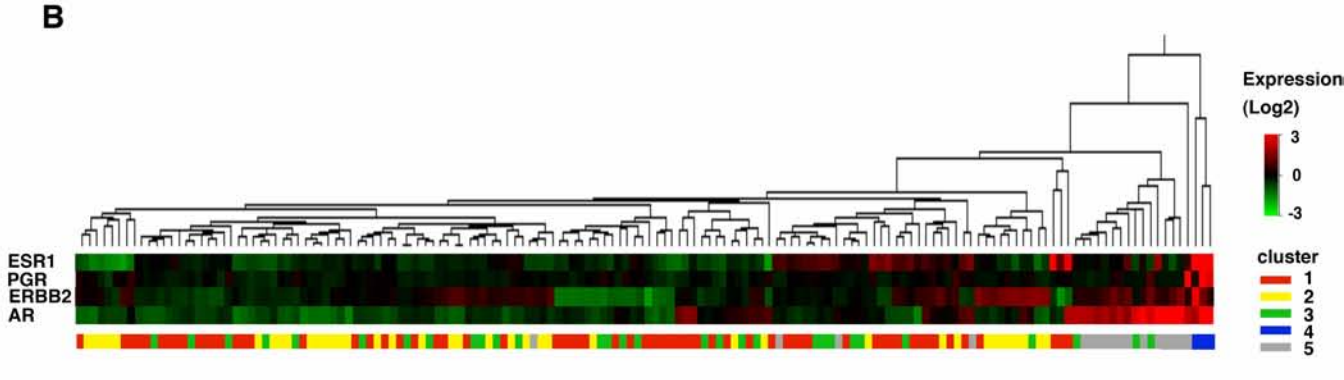


Figure S4. TNBC subtypes identified by IHC. **(A)** Silhouette plot showing the composition (n= number of tumors) and stability (AVG width) of k-means clustering on the TNBC training set. Clusters with a silhouette width, s(i)>0 were considered stable. **(B)** Heatmap displays hierarchical clustering of ERBB2, PGR, ESR1 and AR expression in the tumors identified by IHC. Color bar identifies the cluster associated with each tumor.

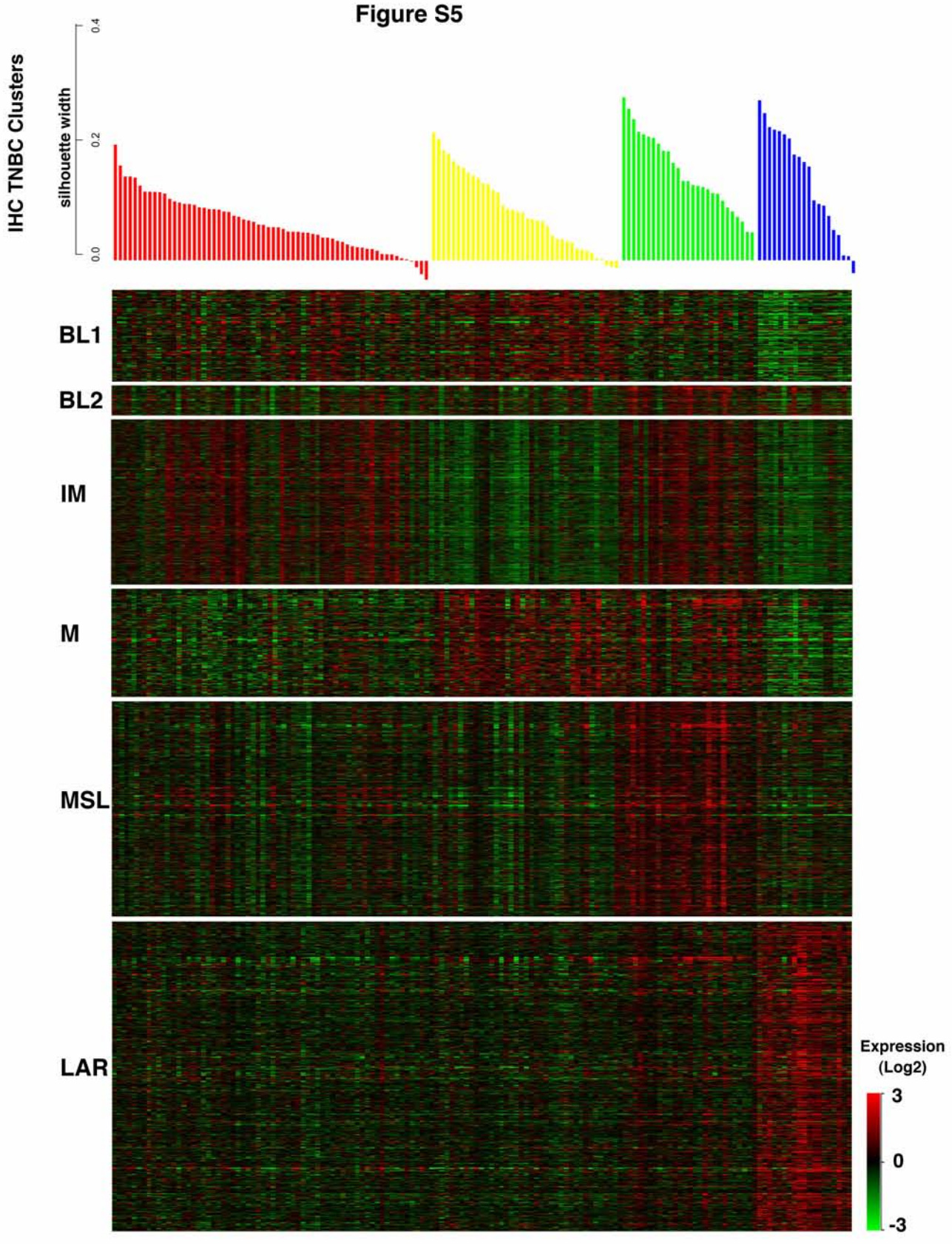


Figure S5. TNBC subtypes identified by IHC display similar GE patterns to the six TNBC subtypes. IHC TNBC clusters are shown in silhouette plot with relative GE for genes unique to the six TNBC subtypes shown below.

Figure S6

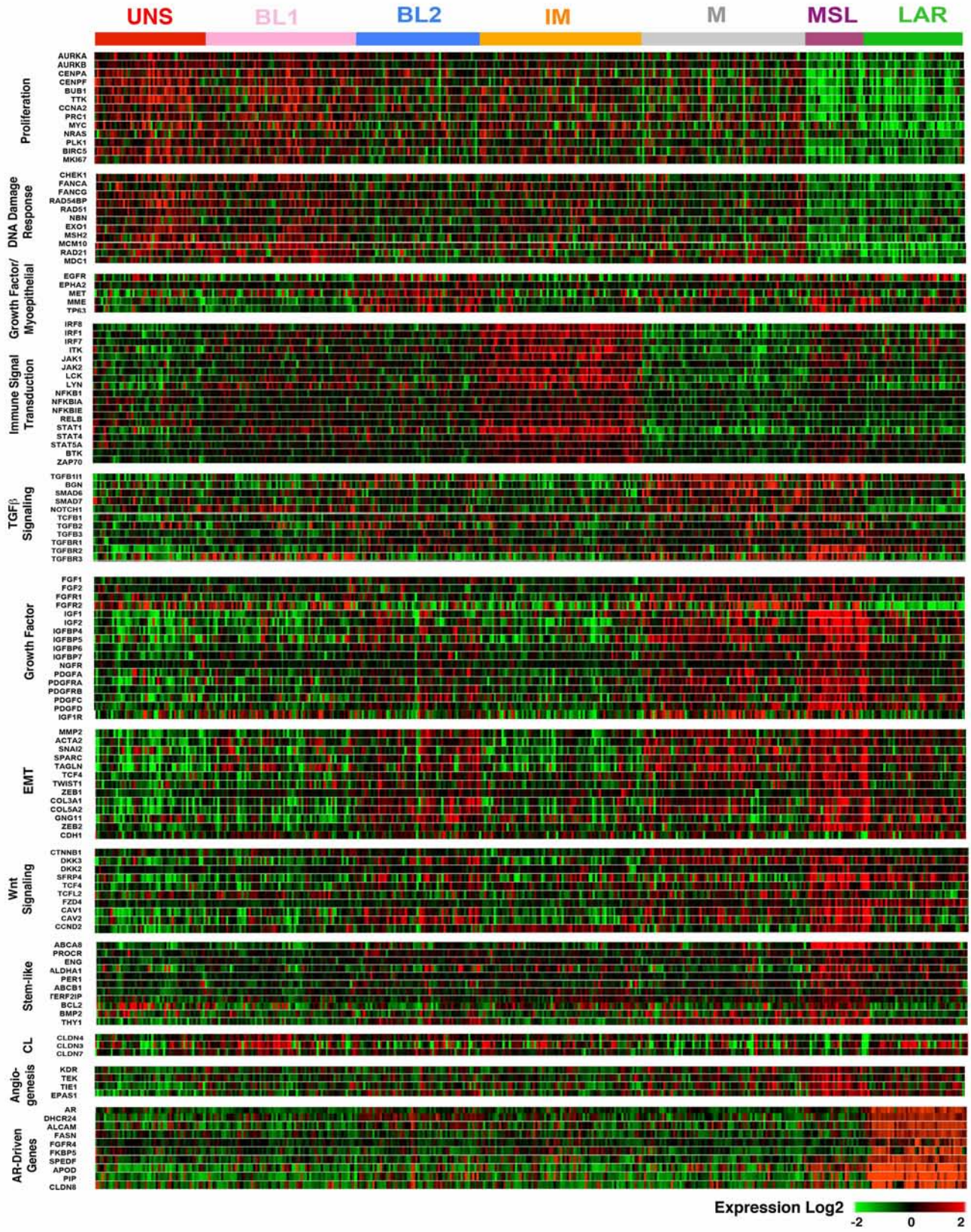


Figure S6. Differential GE across TNBC subtypes. Heatmaps show relative GE (log2, -2 to 2) associated with proliferation, DNA damage response, myoepithelial genes, immune signal transduction, TGFβ signaling, growth factor receptors, EMT, WNT signaling, stem-like, claudin (CL), angiogenesis, AR and of AR target genes across TNBC subtypes (as in Figure 3).

Figure S7

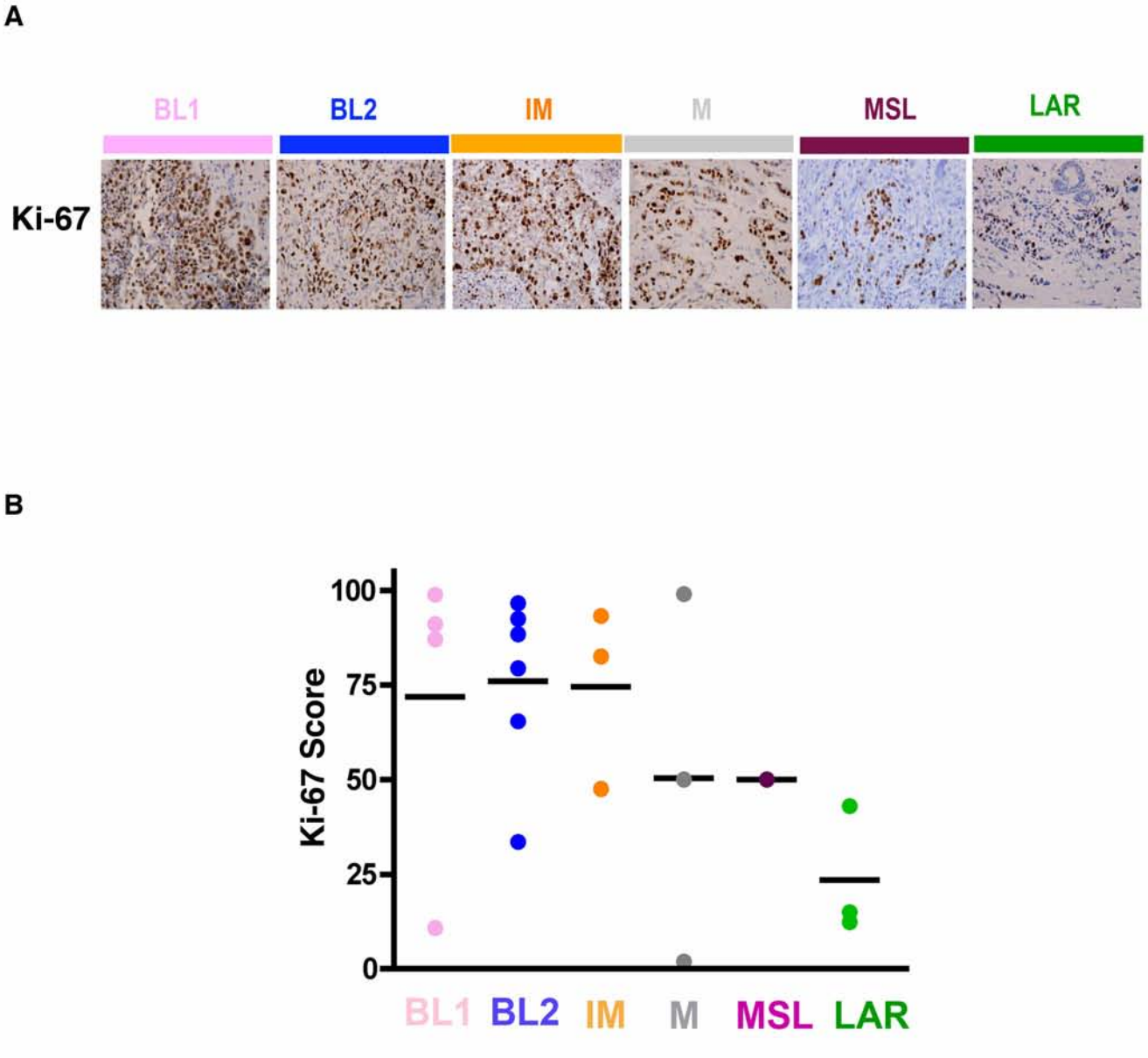


Figure S7. TNBC tumor subtypes differentially stain for the proliferation marker Ki-67. **(A)** Representative micrographs from 20 tumors showing IHC staining of the proliferation marker, Ki-67, in tumors from different TNBC subtypes. **(B)** Dot plot showing the mean and distribution of Ki-67 staining within TNBC subtypes as scored by study pathologist. Subtypes: BL1= basal-like 1, BL2= basal-like 2, IM= immunomodulatory, M= mesenchymal-like, MSL= mesenchymal stem-like and LAR= luminal AR.

Figure 8

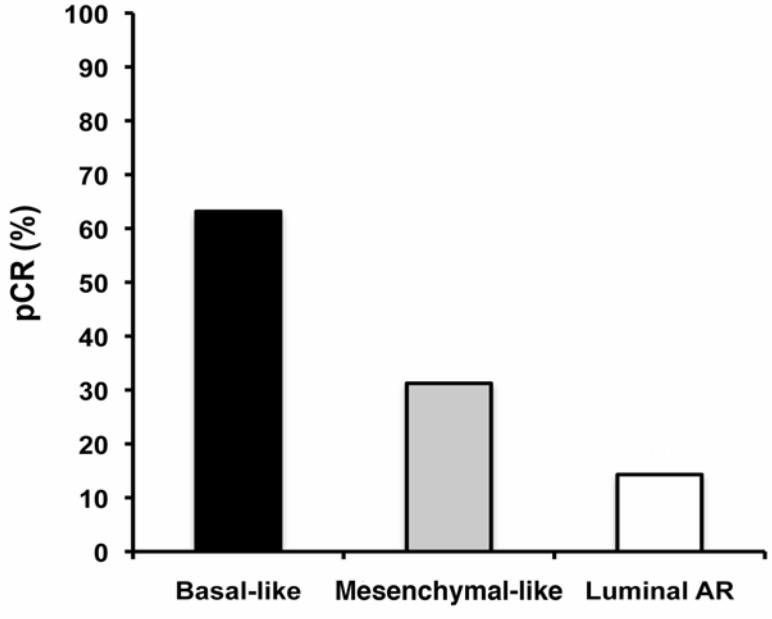


Figure S8. Response rates differ across TNBC subtypes in taxane-treated patients. Percent of patients achieving pathologic complete response (pCR) after taxane-based treatment was significantly ($P = 0.042$, chi-squared analysis) different for patients whose tumors correlated to basal-like (n = 19), mesenchymal-like (n = 16) and luminal AR (n = 7) subtypes, from studies in which response data were available.

Figure S9

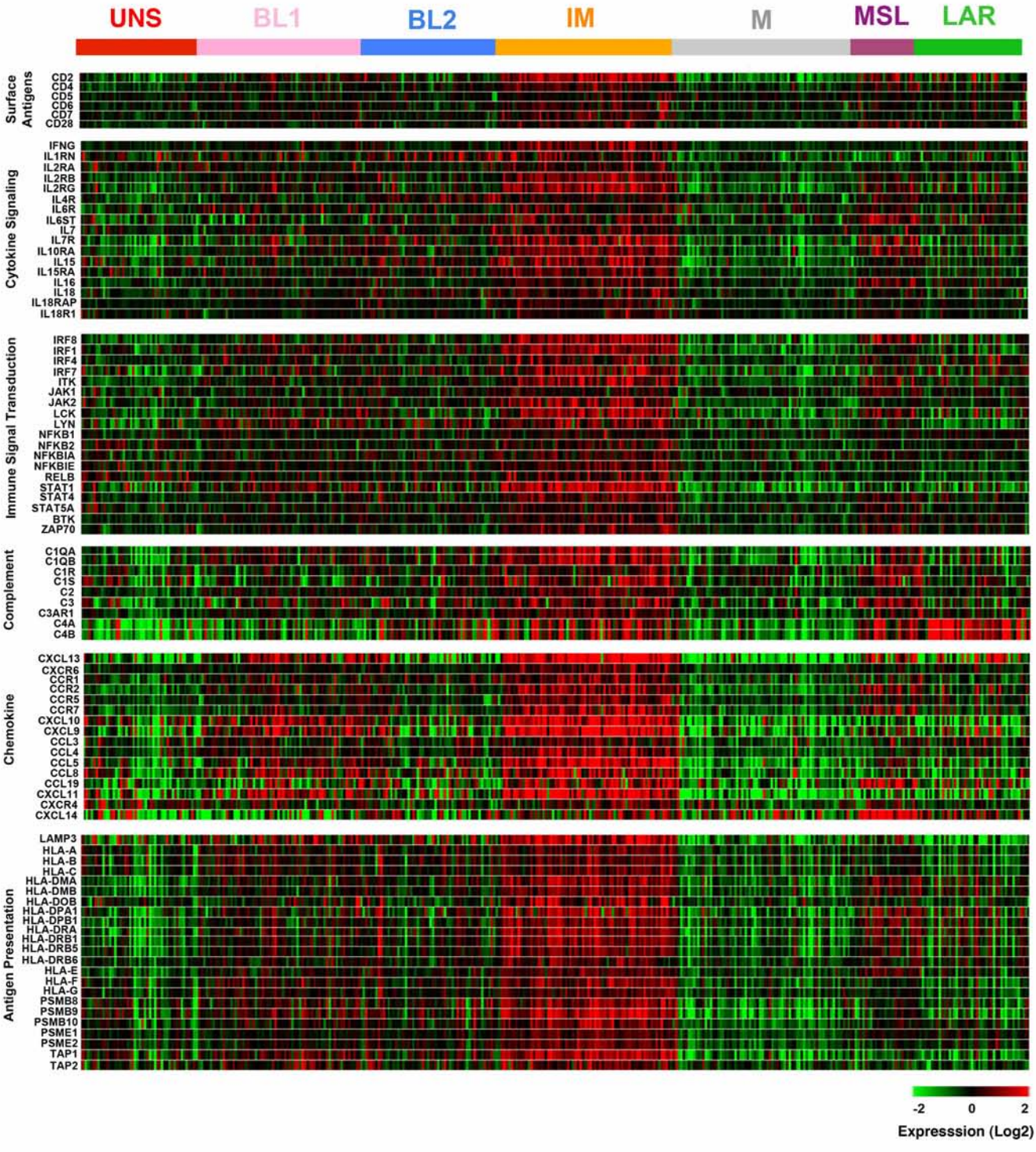


Figure S9. IM subtype is enriched in immune cell signaling. Heatmaps showing the relative GE (log2, -2 to 2) for genes involved in immune cell surface antigens, cytokines, immune cell signal transduction, complement, chemokine, and antigen presentation across TNBC subtypes (as in Figure 3).

Figure S10

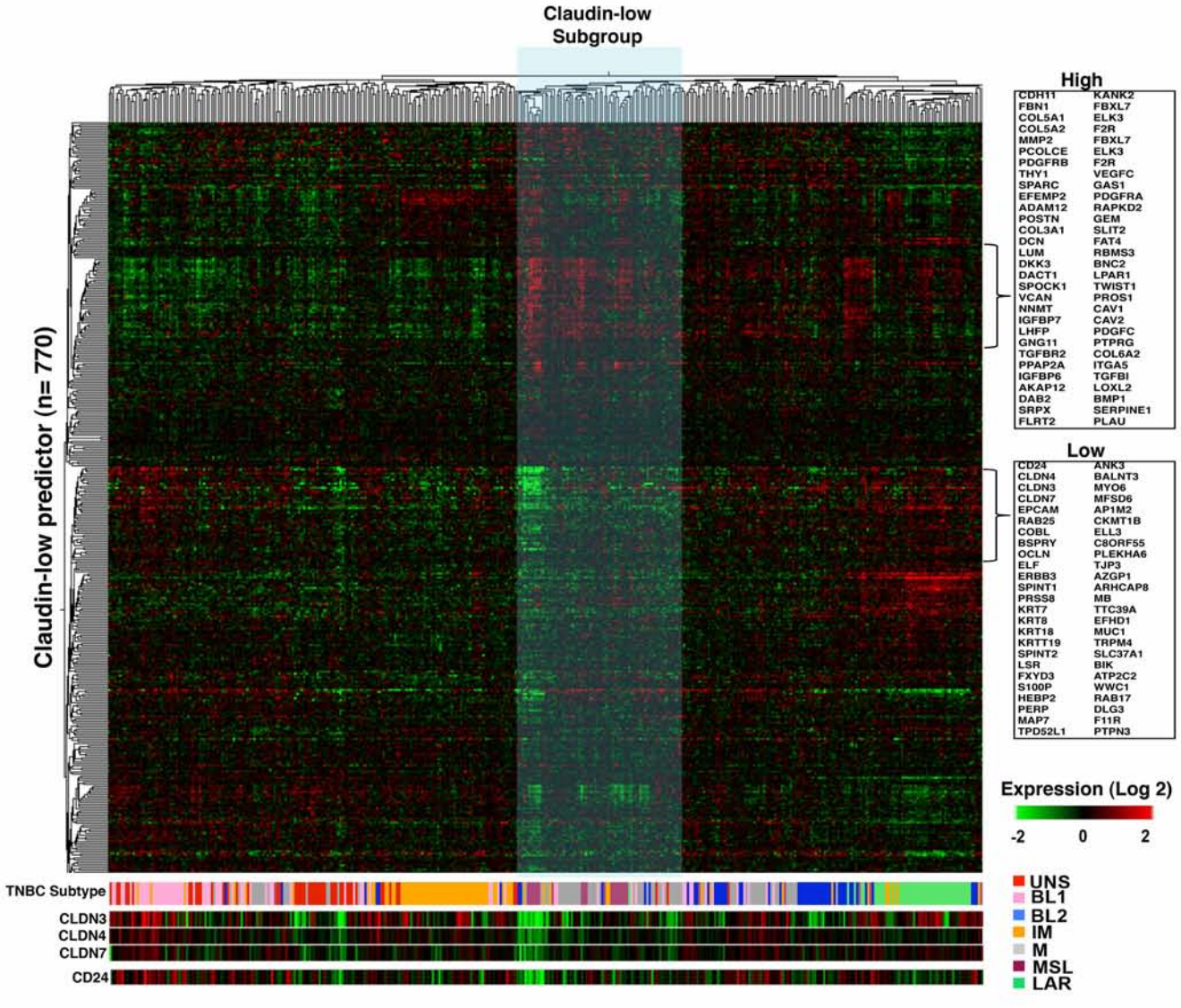


Figure S10. The claudin-low predictor gene set identifies a sub-population of MSL tumors. Unsupervised hierarchical clustering was performed on the training TNBC tumors using genes (n= 770) unique to the claudin-low subgroup [26]. Displayed to the right of the heatmap are the genes that are most differentially expressed (either high or low) in the claudin-low tumor set. Colorbar displays the TNBC subtype and heatmaps show relative levels (Log 2) of claudins (3, 4 and 7) and CD24, markers of this subgroup.

Figure S11

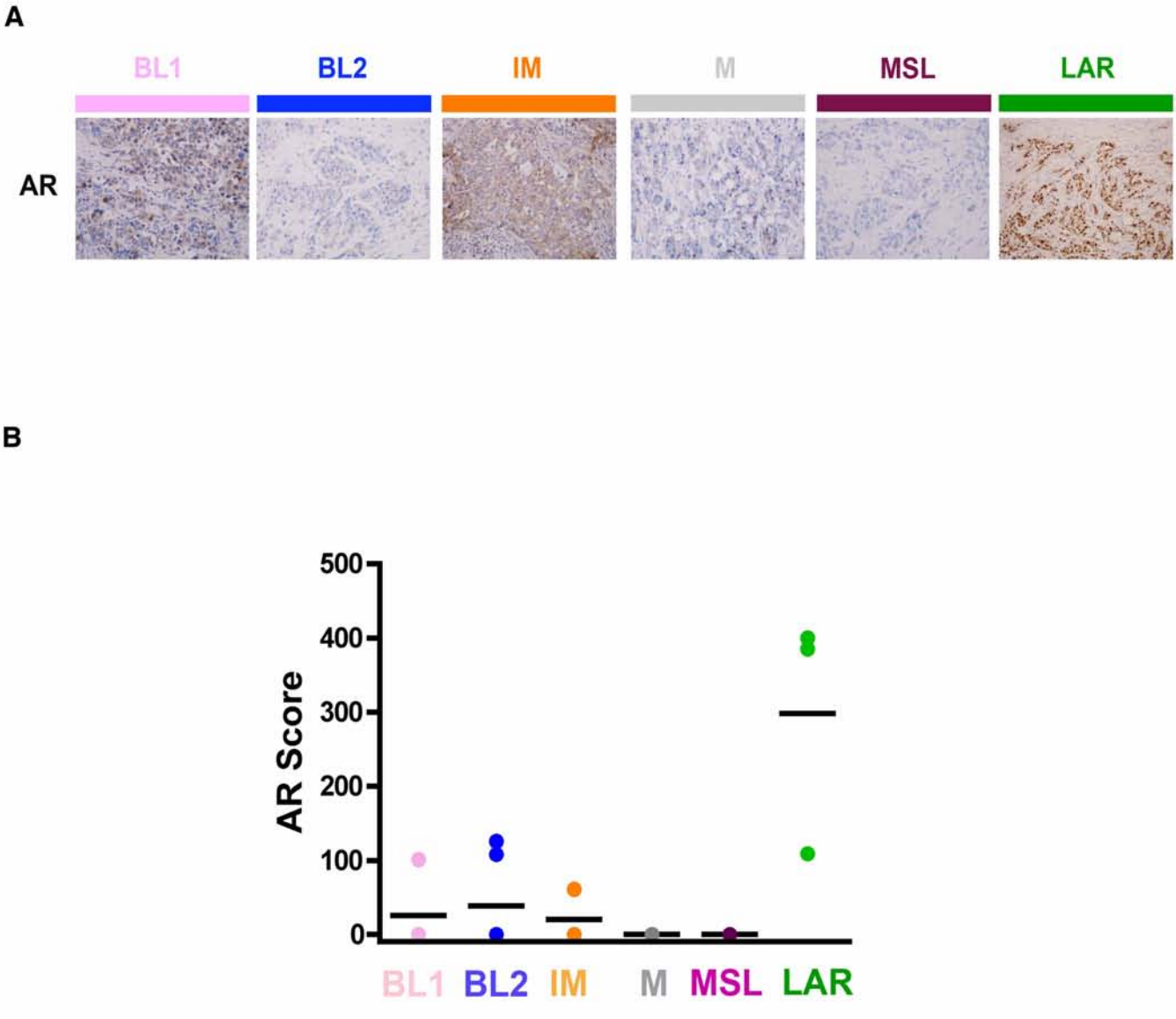


Figure S11. TNBC tumor subtypes differentially stain for AR by IHC. (A) IHC staining for AR from 20 tumors with representative samples from each TNBC subtype shown. **(B)** Dot plot showing the quantification of nuclear AR staining based on intensity and the percent of nuclei staining positive for AR in 20 tumors; Note, in some cases one dot represents overlapping dots from multiple tumors.

Figure S12

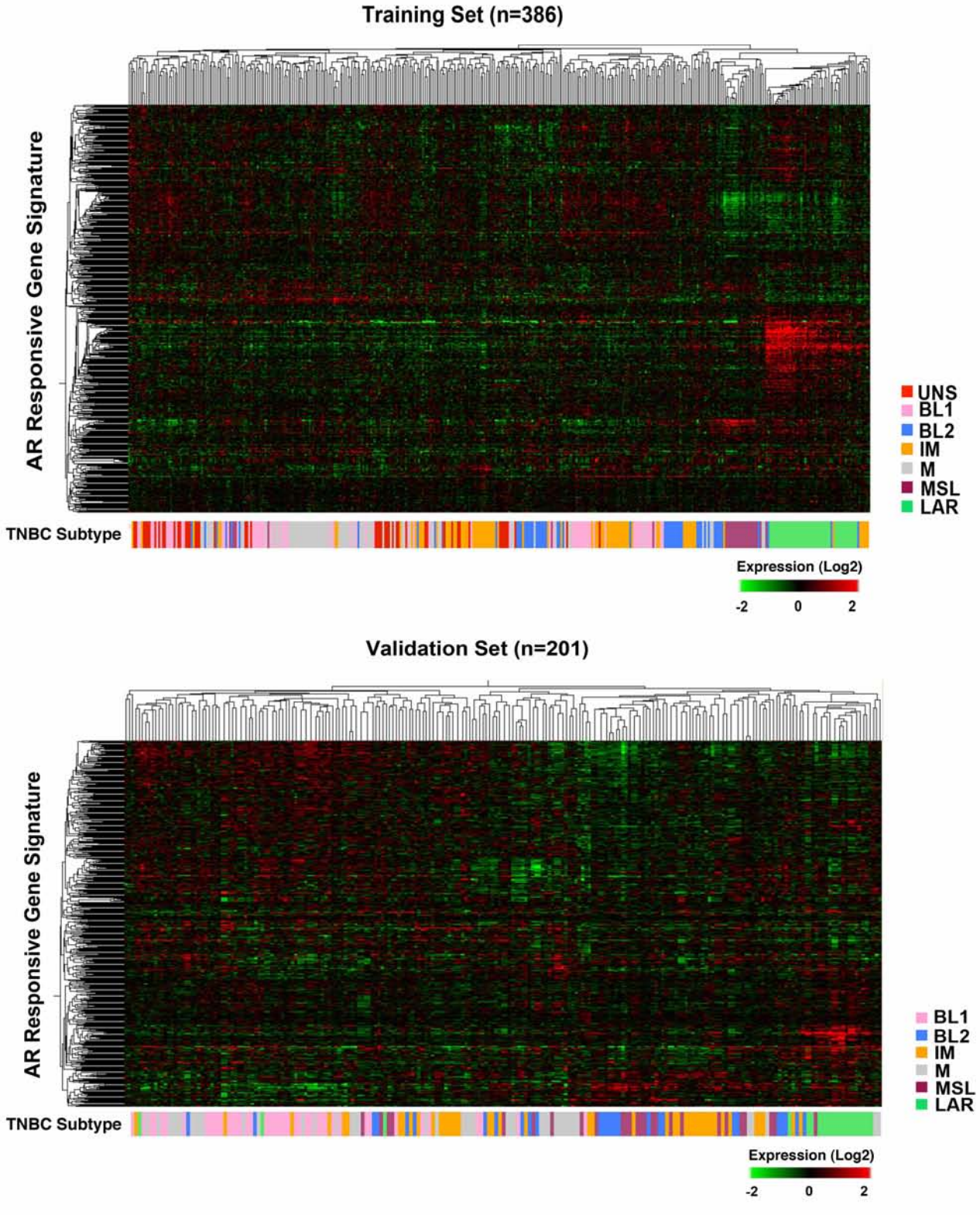


Figure S12. An androgen-inducible gene signature segregates LAR tumors. Hierarchical clustering was performed on both the **(A)** training and **(B)** validation TNBC tumor set using a 559 androgen-inducible gene signature [27].

Figure S13

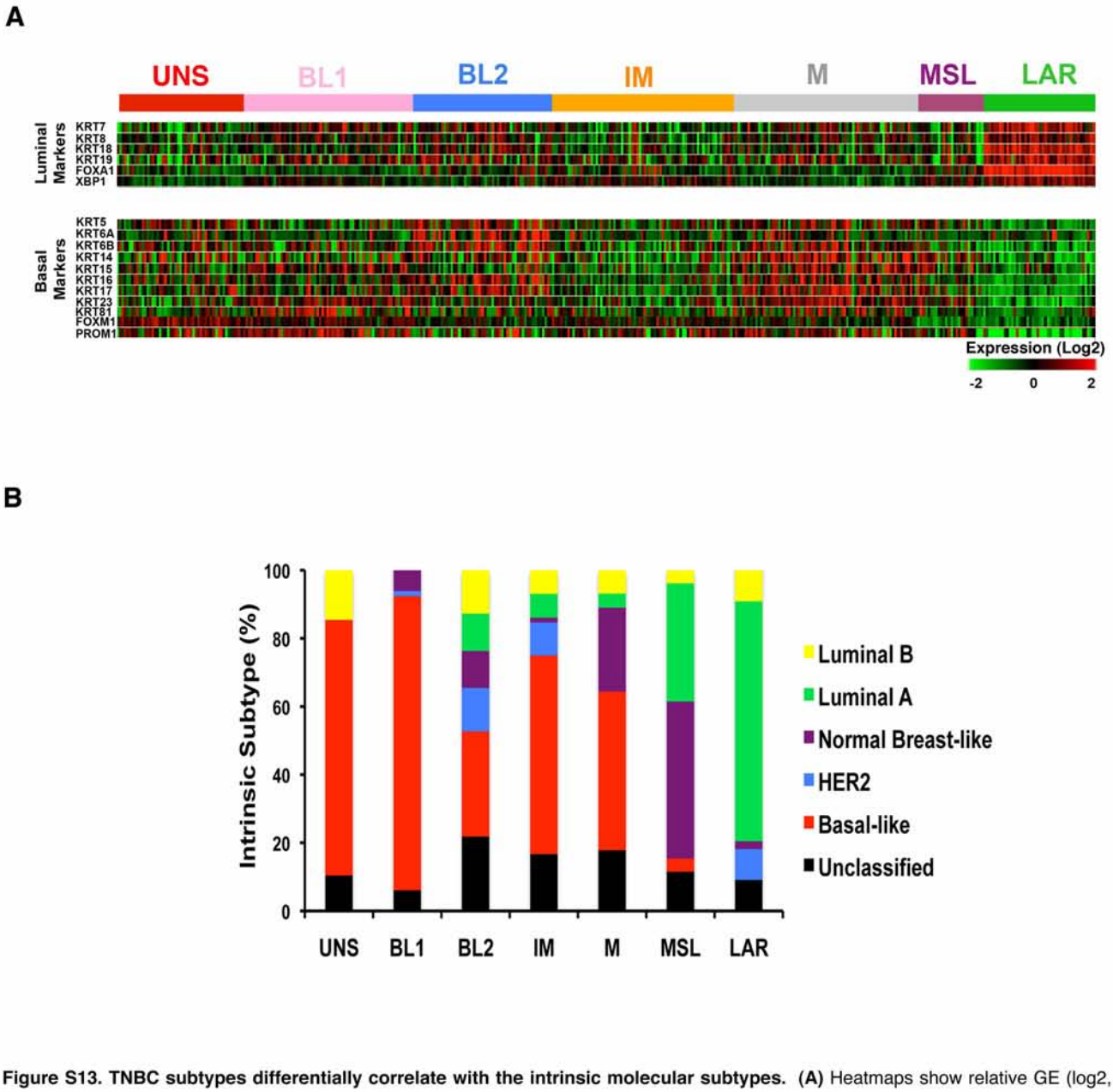
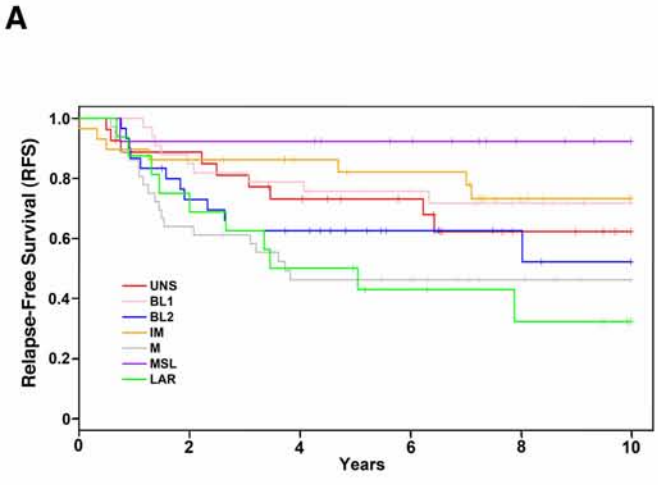
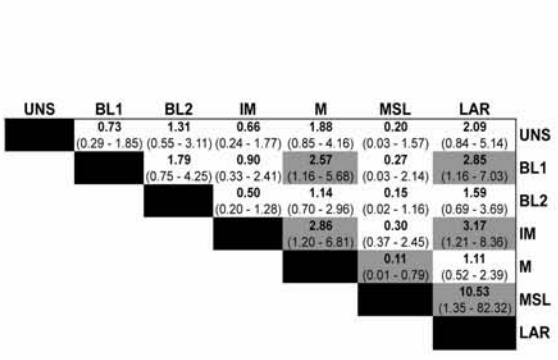
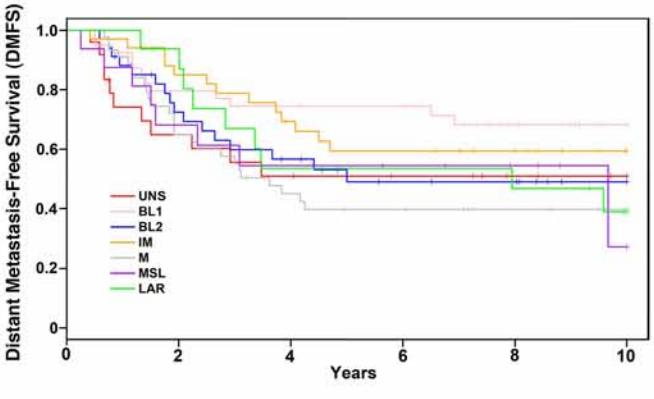


Figure S13. TNBC subtypes differentially correlate with the intrinsic molecular subtypes. **(A)** Heatmaps show relative GE (log2, -2 to 2) of luminal and basal markers of breast cancer across all TNBC subtypes. **(B)** Bar graph shows the distribution of intrinsic molecular subtypes of breast cancer (luminal A and B, normal breast-like, HER2, basal-like or unclassified) within each TNBC subtype (UNS = unstable, BL1= basal-like1, BL2 = basal-like 2, IM = immunomodulatory, M = mesenchymal-like, MSL = mesenchymal stem-like and LAR = luminal AR) as determined by best-fit Spearman correlation to the intrinsic centroids.

Figure S14







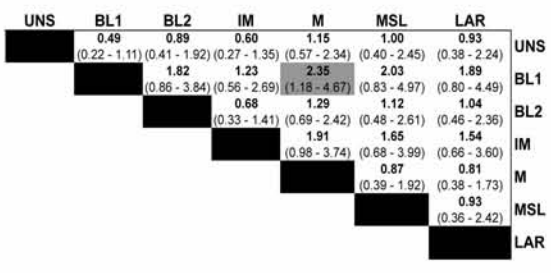
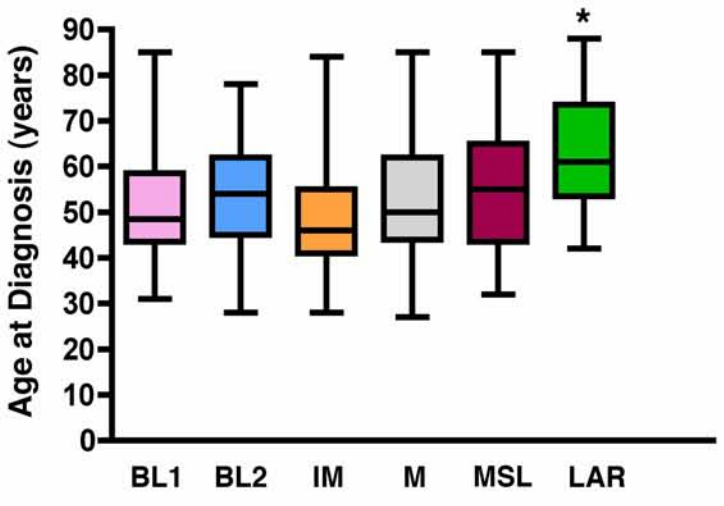


Figure S14. TNBC subtypes differ in relapse-free survival and distant metastasis-free survival. Kaplan-Meier plot showing **(A)** 10-year RFS or **(C)** DMFS in TNBC subtypes. RFS **(B)** and DMFS **(D)** hazard ratios (bold) and 95% CI (parentheses) for patients from TNBC subtypes. Shaded boxes indicate significant ($P < 0.05$) comparisons.

Figure S15

A



B

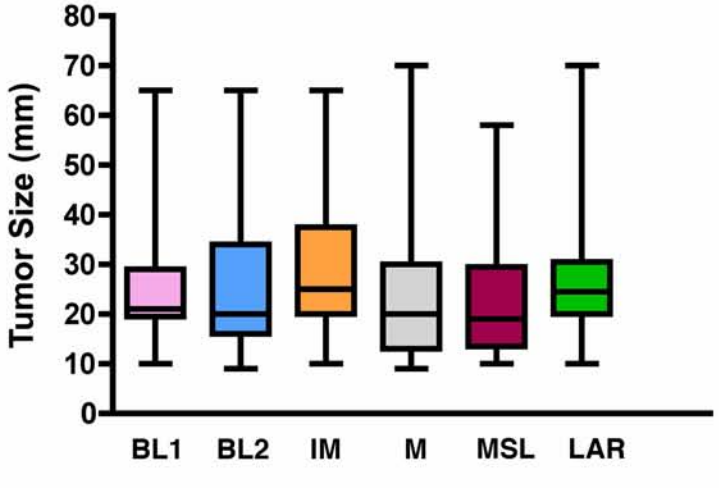
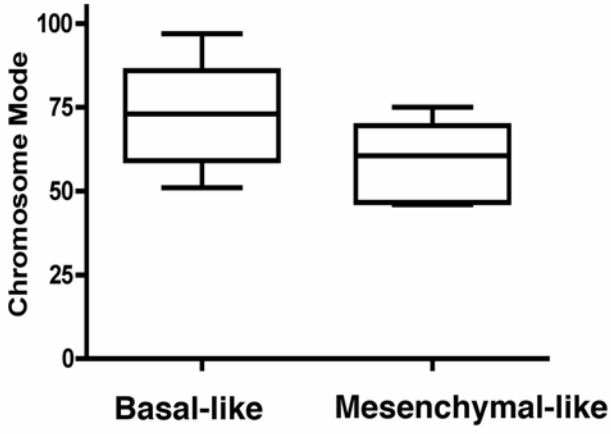


Figure S15. TNBC subtypes differ in age, but are similar size upon diagnosis. Box plots show the median (horizontal line), range (rectangle) and SD (error bars) of **(A)** age at diagnosis and **(B)** tumor size (mm) between TNBC subtypes (as in Figure 3). * $P = 9.0e\text{-}6$

Figure S16

A



B

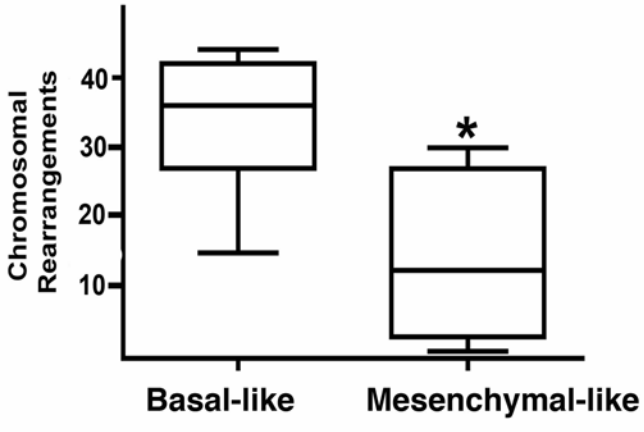


Figure S16. Chromosomal Aberrations in TNBC cell lines. **(A)** Box plots depicting the average number of chromosomes (mode) in breast cancer cell lines correlating to basal-like (n=8) vs. mesenchymal-like (n=6) subtypes. **(B)** Box plots depicting the average number of chromosomal rearrangements (translocations, inversions, and deletions) in basal-like vs. mesenchymal-like subtypes. Chromosome mode and rearrangements were obtained from the Departments of Pathology and Oncology, University of Cambridge (http://www.path.cam.ac.uk/~pawefish/cell%20line%20catalogues/breast-cell-lines.htm). *P <0.01 by unpaired t-test

Figure S17

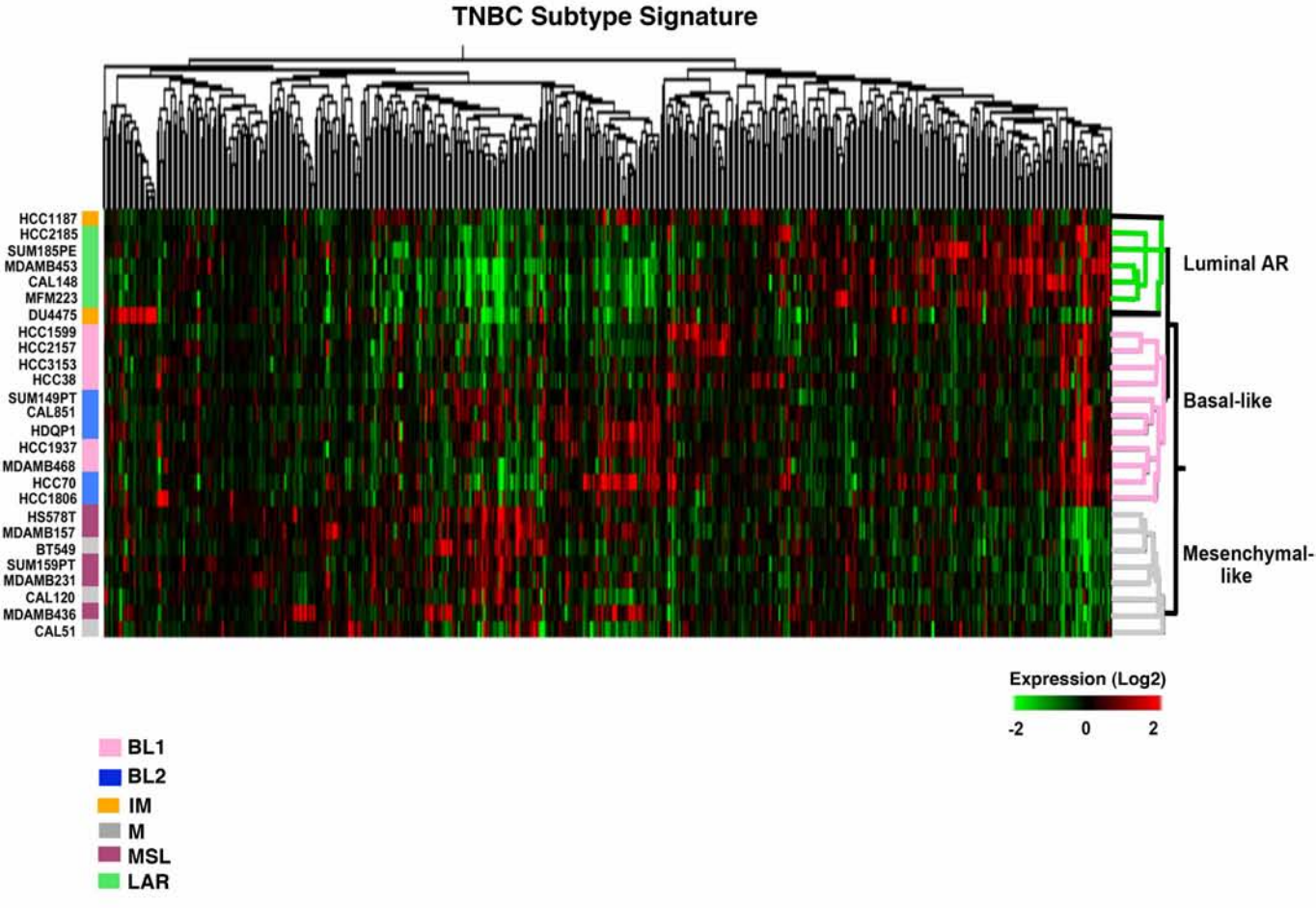
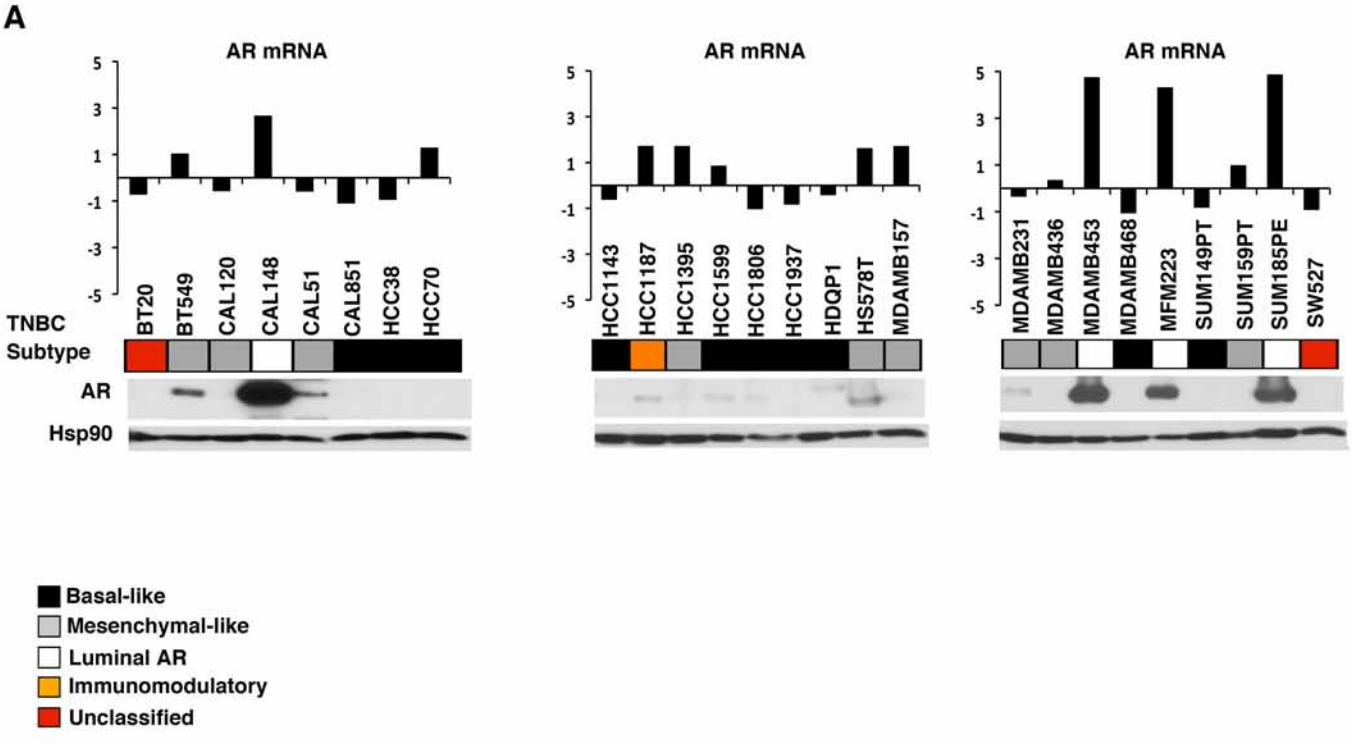
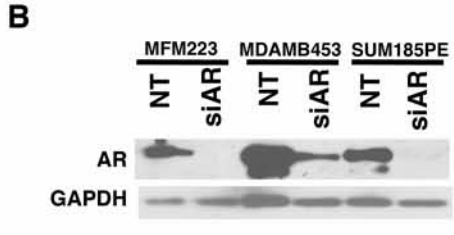


Figure S17. TNBC cell lines cluster in three major groups: luminal AR, basal-like and mesenchymal-like. Unsupervised hierarchical clustering of TNBC cell lines performed on genes unique to TNBC subtypes (n= 2188). Colorbar shows the best correlation of cell lines to the TNBC subtypes.

Figure S18





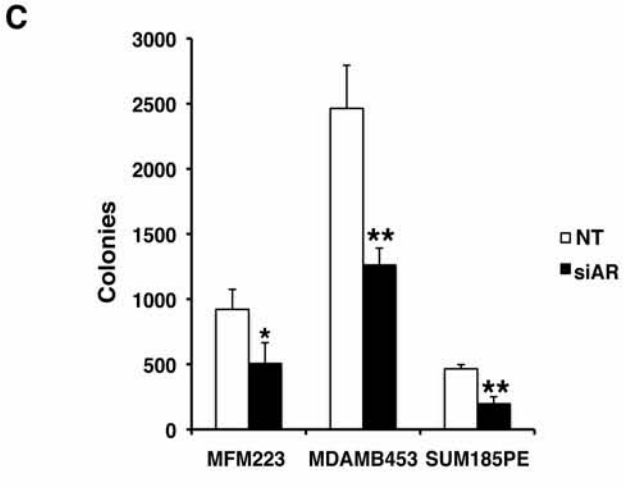


Figure S18. LAR cell lines depend on AR expression for colony formation. **(A)** Top panel depicts relative AR mRNA levels obtained from GE microarrays performed on TNBC cell lines (log 2, centered on 0). Color bar identifies TNBC subtype classification for each cell line (immunomodulatory shown in orange and unclassified shown in red). Immunoblots showing relative expression AR protein in TNBC cell lines, Hsp90 serves as a loading control. **(B)** AR expression 72h following transfection with siRNA pools of non-targeting (NT) or targeting AR in MFM-223, MDA-MB-453 and SUM185PE cells. GAPDH expression serves as a loading control. **(C)** Colony formation of MFM-223, MDA-MB-453 and SUM185PE cells 14d following siRNA transfection of AR or non-targeting control (NT).

Table S7. Top gene ontologies for TNBC subtypes in the training and validation datasets

Subtype	Training Set (386)	Validation Set (201)
BL1	**DNA REPLICATION REACTOME** **CELL CYCLE KEGG** **RNA POLYMERASE** **CELL CYCLE** **GLYCOSPHINGOLIPID BIOSYNTHESIS NEO LACTOSERIES** **ATRBRCAPATHWAY** UBIQUITIN MEDIATED PROTEOLYSIS **G2 PATHWAY** PROTEASOME **G1 TO S CELL CYCLE REACTOME** **METHIONINE METABOLISM** **FASPATHWAY** **CHOLESTEROL BIOSYNTHESIS** **PYRIMIDINE METABOLISM** **AMINOACYL TRNA BIOSYNTHESIS** DNA POLYMERASE CARBON FIXATION **ALANINE AND ASPARTATE METABOLISM** PROTEASOME **CELLCYCLEPATHWAY** CDMACPATHWAY KREBS TCA CYCLE	**CELL CYCLE KEGG** **DNA REPLICATION REACTOME** **G2 PATHWAY** **METHIONINE METABOLISM** **CHOLESTEROL BIOSYNTHESIS** **CELL CYCLE** **RNA POLYMERASE** **ATRBRCAPATHWAY** **G1 TO S CELL CYCLE REACTOME** CASPASE CASCADE GLUTAMATE METABOLISM **AMINOACYL TRNA BIOSYNTHESIS** **GLYCOSPHINGOLIPID BIOSYNTHESIS NEO LACTOSERIES** G1 AND S PHASES CYSTEINE METABOLISM ATMPATHWAY MRNA PROCESSING REACTOME BIOSYNTHESIS OF STEROIDS KREBS TCA CYCLE **PYRIMIDINE METABOLISM** PROTEASOMEPATHWAY SELENOAMINO ACID METABOLISM **CELLCYCLEPATHWAY** **ALANINE AND ASPARTATE METABOLISM** GAQ PATHWAY AMINOACYL TRNA BIOSYNTHESIS **FASPATHWAY** **PROTEASOME**
BL2	PORPHYRIN AND CHLOROPHYLL METABOLISM **GATA3PATHWAY** STEMPATHWAY PARKINSONS DISEASE STARCH AND SUCROSE METABOLISM EGFPATHWAY INSULINPATHWAY RENAL CELL CARCINOMA WNT SIGNALING AMINOACYL TRNA BIOSYNTHESIS GALACTOSE METABOLISM O GLYCAN BIOSYNTHESIS ATMPATHWAY HUNTINGTONS DISEASE IL6PATHWAY G1 TO S CELL CYCLE REACTOME FOCAL ADHESION IGF1PATHWAY UBIQUITIN MEDIATED PROTEOLYSIS GLIOMA TPOPATHWAY G2PATHWAY **WNT BETA CATENIN PATHWAY** INSULIN RECEPTOR PATHWAY IN CARDIAC MYOCYTES **METPATHWAY** P53 SIGNALING PATHWAY **CARDIACEGFPATHWAY** RASPATHWAY **NFKBPATHWAY** UCALPAINPATHWAY GSK3PATHWAY **GLYCOLYSIS AND GLUCONEOGENESIS** EPOPATHWAY **SPRYPATHWAY**	BLADDER CANCER **GATA3PATHWAY** EPITHELIAL CELL SIGNALING IN HELICOBACTER PYLORI INFECTION **CARDIACEGFPATHWAY** NICOTINATE AND NICOTINAMIDE METABOLISM EDG1PATHWAY PANTOTHENATE AND COA BIOSYNTHESIS TOLLPATHWAY RENIN ANGIOTENSIN SYSTEM ECM RECEPTOR INTERACTION PHOTOSYNTHESIS NTHIPATHWAY AMINOSUGARS METABOLISM EICOSANOID SYNTHESIS DORSO VENTRAL AXIS FORMATION **NGFPATHWAY** **SPRYPATHWAY** **ECMPATHWAY** **METPATHWAY** CDMACPATHWAY ARFPATHWAY **P38 MAPK PATHWAY** CHOLERA INFECTION HYPERTROPHY MODEL **WNT BETA CATENIN PATHWAY** **ERKPATHWAY** FMLPPATHWAY **GLYCOLYSIS** ATMPATHWAY **IGF1RPATHWAY** GLYCOSAMINOGLYCAN DEGRADATION **GALACTOSE METABOLISM** **ERK1 ERK2 MAPK PATHWAY** **GLUCONEOGENESIS**

	TRKA RECEPTOR PHOSPHOINOSITIDE 3 KINASE PATHWAY PROTEASOME **P38 MAPK PATHWAY** PATHOGENIC ESCHERICHIA COLI INFECTION EPEC CHOLERA INFECTION TIDPATHWAY **ECMPATHWAY** PDGFPATHWAY ARGININE AND PROLINE METABOLISM **GLYCOLYSIS** PROTEASOMEPATHWAY NTHIPATHWAY EPITHELIAL CELL SIGNALING IN HELICOBACTER PYLORI INFECTION **NGFPATHWAY** **ETSPATHWAY** **ERK1 ERK2 MAPK PATHWAY** **GALACTOSE METABOLISM** GAP JUNCTION ERYTHPATHWAY **GLYCOLYSIS AND GLUCONEOGENESIS** MYOCYTE AD PATHWAY	1 AND 2 METHYLNAPHTHALENE DEGRADATION VEGFPATHWAY **ETSPATHWAY** PROSTAGLANDIN SYNTHESIS REGULATION BENZOATE DEGRADATION VIA COA LIGATION **NFKBPATHWAY** HEPARAN SULFATE BIOSYNTHESIS
IM	**CTLA4PATHWAY** **NO2IL12PATHWAY** **TH1TH2PATHWAY** **CSKPATHWAY** **NKTPATHWAY** **COMPPATHWAY** **IL7PATHWAY** **PROTEASOME** **NKCELLSPATHWAY** **TYPE I DIABETES MELLITUS** **AMIPATHWAY** **ANTIGEN PROCESSING AND PRESENTATION** **IL12PATHWAY** **LAIRPATHWAY** **TNFR2PATHWAY** **DCPATHWAY** **TIDPATHWAY** **APOPTOSIS GENMAPP** **NFKBPATHWAY** **PROTEASOMEPATHWAY** **T CELL SIGNAL TRANSDUCTION** **TUMOR NECROSIS FACTOR PATHWAY** **CASPASEPATHWAY** **APOPTOSIS** **NATURAL KILLER CELL MEDIATED CYTOTOXICITY** **TOB1PATHWAY** **B CELL RECEPTOR SIGNALING PATHWAY** **T CELL RECEPTOR SIGNALING PATHWAY** **CYTOKINEPATHWAY** **CELL ADHESION MOLECULES** **DEATHPATHWAY** **APOPTOSIS KEGG** **TOLL LIKE RECEPTOR SIGNALING PATHWAY** **PROTEASOME** **HEMATOPOIETIC CELL LINEAGE** **RELAPATHWAY** **IL2PATHWAY** **MITOCHONDRIAPATHWAY** **TOLLPATHWAY** **B CELL ANTIGEN RECEPTOR** **INFLAMPATHWAY** **IL2RBPATHWAY** **IL3PATHWAY** **STEMPATHWAY** **BCR SIGNALING PATHWAY** **HIVNEFPATHWAY** **TCRPATHWAY**	**NO2IL12PATHWAY** **CTLA4PATHWAY** **IL12PATHWAY** **TYPE I DIABETES MELLITUS** **NKCELLSPATHWAY** **TH1TH2PATHWAY** **TNFR2PATHWAY** **CSKPATHWAY** **IL7PATHWAY** **AMIPATHWAY** **ANTIGEN PROCESSING AND PRESENTATION** **NFKBPATHWAY** **TUMOR NECROSIS FACTOR PATHWAY** **PROTEASOME** **NKTPATHWAY** **T CELL SIGNAL TRANSDUCTION** **DEATHPATHWAY** **APOPTOSIS GENMAPP** **DCPATHWAY** **RELAPATHWAY** **TOB1PATHWAY** **APOPTOSIS** **APOPTOSIS KEGG** **T CELL RECEPTOR SIGNALING PATHWAY** **CYTOKINEPATHWAY** **CASPASEPATHWAY** **TIDPATHWAY** **IL2PATHWAY** **NATURAL KILLER CELL MEDIATED CYTOTOXICITY** **CASPASE CASCADE** **LAIRPATHWAY** **HEMATOPOIETIC CELL LINEAGE** **IL3PATHWAY** **COMPPATHWAY** **TCRPATHWAY** **B CELL RECEPTOR SIGNALING PATHWAY** **STEMPATHWAY** **MITOCHONDRIAPATHWAY** **CELL ADHESION MOLECULES** **TOLLPATHWAY** **STRESSPATHWAY** **IL2RBPATHWAY** **INFLAMPATHWAY** **CYTOKINE CYTOKINE RECEPTOR INTERACTION** **PIP3 SIGNALING IN B LYMPHOCYTES** **TOLL LIKE RECEPTOR SIGNALING PATHWAY** **41BBPATHWAY**

	CASPASE CASCADE **PIP3 SIGNALING IN B LYMPHOCYTES** **NTHIPATHWAY** **41BBPATHWAY** **CCR5PATHWAY** **CYTOKINE CYTOKINE RECEPTOR INTERACTION** **INTERLEUKIN 4 PATHWAY** **STRESSPATHWAY** **IL1RPATHWAY** **APOPTOSIS** **EICOSANOID SYNTHESIS** UBIQUITIN MEDIATED PROTEOLYSIS **CELLCYCLEPATHWAY** **CALCINEURIN NF AT SIGNALING** **FC EPSILON RI SIGNALING PATHWAY** **AMINOACYL TRNA BIOSYNTHESIS** **JAK STAT SIGNALING PATHWAY** **FASPATHWAY** **PTEN PATHWAY** **B CELL RECEPTOR COMPLEXES** DNA REPLICATION REACTOME **FOLATE BIOSYNTHESIS** **ATRBRCAPATHWAY** **IL6PATHWAY** **LEUKOCYTE TRANSENDOTHELIAL MIGRATION** **GLEEVECPATHWAY** **TNFR1PATHWAY** COMPLEMENT AND COAGULATION CASCADES **BCRPATHWAY** **EPOPATHWAY** **GHPATHWAY** **TPOPATHWAY** AMINOACYL TRNA BIOSYNTHESIS **PEPTIDE GPCRS** N GLYCAN BIOSYNTHESIS ALZHEIMERS DISEASE **ACUTE MYELOID LEUKEMIA** GAQ PATHWAY **CD40PATHWAYMAP** METHIONINE METABOLISM GPCRDB CLASS B SECRETIN LIKE **IL4RECEPTOR IN B LYPHOCYTES** **CERAMIDEPATHWAY** FCER1PATHWAY G1 TO S CELL CYCLE REACTOME PARKINSONS DISEASE CHEMICALPATHWAY CELL CYCLE KEGG SNARE INTERACTIONS IN VESICULAR TRANSPORT ATMPATHWAY NICOTINATE AND NICOTINAMIDE METABOLISM KERATINOCYTEPATHWAY HYPERTROPHY MODEL **ADIPOCYTOKINE SIGNALING PATHWAY** **DENTATORUBROPALLIDOLUYSIAN ATROPHY** FMLPPATHWAY DICTYOSTELIUM DISCOIDEUM CAMP CHEMOTAXIS PATHWAY BIOPEPTIDESPATHWAY PANCREATIC CANCER FAS SIGNALING PATHWAY RACCYCDPATHWAY AMINOSUGARS METABOLISM	**B CELL ANTIGEN RECEPTOR** **BCR SIGNALING PATHWAY** **PROTEASOME** **CERAMIDEPATHWAY** **APOPTOSIS** **JAK STAT SIGNALING PATHWAY** **EPOPATHWAY** **NTHIPATHWAY** **CD40PATHWAYMAP** **CALCINEURIN NF AT SIGNALING** **HIVNEFPATHWAY** **EICOSANOID SYNTHESIS** **BCRPATHWAY** **FC EPSILON RI SIGNALING PATHWAY** **KERATINOCYTEPATHWAY** **AMINOACYL TRNA BIOSYNTHESIS** **CELLCYCLEPATHWAY** **B CELL RECEPTOR COMPLEXES** **IL6PATHWAY** PROTEASOMEPATHWAY **CCR5PATHWAY** **ATRBRCAPATHWAY** **GHPATHWAY** **FOLATE BIOSYNTHESIS** **PEPTIDE GPCRS** **GLEEVECPATHWAY** **FASPATHWAY** ABC TRANSPORTERS GENERAL **INTERLEUKIN 4 PATHWAY** **IL4RECEPTOR IN B LYPHOCYTES** **ADIPOCYTOKINE SIGNALING PATHWAY** PROSTAGLANDIN AND LEUKOTRIENE METABOLISM CDMACPATHWAY **TPOPATHWAY** HSP27PATHWAY **TNFR1PATHWAY** INOSITOL PHOSPHATE METABOLISM **ACUTE MYELOID LEUKEMIA** **LEUKOCYTE TRANSENDOTHELIAL MIGRATION** **IL1RPATHWAY** **DENTATORUBROPALLIDOLUYSIAN ATROPHY** OVARIAN INFERTILITY GENES **PTEN PATHWAY**
M	CELL COMMUNICATION HEPARAN SULFATE BIOSYNTHESIS ECM RECEPTOR INTERACTION **ALKPATHWAY** **UCALPAINPATHWAY** STRIATED MUSCLE CONTRACTION BASAL CELL CARCINOMA	PTENPATHWAY DNA REPLICATION REACTOME RIBOSOME TRANSLATION FACTORS RNA POLYMERASE DNA POLYMERASE BASAL TRANSCRIPTION FACTORS

	REGULATION OF THE ACTIN CYTOSKELETON BY RHO GTPASES TGF BETA SIGNALING PATHWAY HDACPATHWAY HEDGEHOG SIGNALING PATHWAY INOSITOL PHOSPHATE METABOLISM IGF1PATHWAY **ECMPATHWAY** ERK5PATHWAY FOCAL ADHESION EDG1PATHWAY CARM ERPATHWAY INSULINPATHWAY ETHER LIPID METABOLISM **TELPATHWAY** NOTCH SIGNALING PATHWAY WNT BETA CATENIN PATHWAY MEF2DPATHWAY INTEGRINPATHWAY MELANOGENESIS VEGFPATHWAY **WNTPATHWAY** ETSPATHWAY DORSO VENTRAL AXIS FORMATION	IGF1MTORPATHWAY RNA TRANSCRIPTION REACTOME CHOLESTEROL BIOSYNTHESIS GLYCOSYLPHOSPHATIDYLINOSITOL ANCHOR BIOSYNTHESIS **ECMPATHWAY** **UCALPAINPATHWAY** EIF4PATHWAY RIBOSOMAL PROTEINS CELL CYCLE KEGG MTORPATHWAY **REGULATION OF THE ACTIN CYTOSKELETON BY RHO GTPASES** G1 TO S CELL CYCLE REACTOME MRNA PROCESSING REACTOME UBIQUITIN MEDIATED PROTEOLYSIS **WNTPATHWAY** **ALKPATHWAY** G2PATHWAY **TELPATHWAY** CELL CYCLE ATRBRCAPATHWAY
MSL	**PROSTAGLANDIN SYNTHESIS REGULATION** **BADPATHWAY** **LAIRPATHWAY** **IL7PATHWAY** **COMPPATHWAY** **RENIN ANGIOTENSIN SYSTEM** **AMIPATHWAY** **CSKPATHWAY** **ERYTHPATHWAY** **ECM RECEPTOR INTERACTION** **TOB1PATHWAY** **COMPLEMENT AND COAGULATION CASCADES** **CARDIACEGFPATHWAY** **HISTIDINE METABOLISM** **GATA3PATHWAY** **NDKDYNAMINPATHWAY** **CCR5PATHWAY** FOCAL ADHESION **BETA ALANINE METABOLISM** I**NTRINSICPATHWAY** **PAR1PATHWAY** **ALKPATHWAY** **CALCINEURINPATHWAY** NTHIPATHWAY **CTLA4PATHWAY** BETA ALANINE METABOLISM **GCRPATHWAY** **HEMATOPOIETIC CELL LINEAGE** **NO1PATHWAY** **VIPPATHWAY** UCALPAINPATHWAY EDG1PATHWAY **IL3PATHWAY** IGF1PATHWAY TCRPATHWAY EICOSANOID SYNTHESIS NKCELLSPATHWAY VALINE LEUCINE AND ISOLEUCINE DEGRADATION PPARAPATHWAY SPRYPATHWAY HDACPATHWAY STATIN PATHWAY PHARMGKB GLYCAN STRUCTURES DEGRADATION SMOOTH MUSCLE CONTRACTION **TH1TH2PATHWAY** ECMPATHWAY	**NO2IL12PATHWAY** **CSKPATHWAY** **LAIRPATHWAY** **TOB1PATHWAY** **CTLA4PATHWAY** **AMIPATHWAY** **TH1TH2PATHWAY** **RENIN ANGIOTENSIN SYSTEM** **PAR1PATHWAY** NKTPATHWAY **IL7PATHWAY** NKCELLSPATHWAY DCPATHWAY **ERYTHPATHWAY** **HEMATOPOIETIC CELL LINEAGE** TYPE I DIABETES MELLITUS **IL3PATHWAY** GLYCOSAMINOGLYCAN DEGRADATION IL12PATHWAY **IL2PATHWAY** **PROSTAGLANDIN SYNTHESIS REGULATION** EICOSANOID SYNTHESIS **COMPPATHWAY** **CCR5PATHWAY** GLYCAN STRUCTURES DEGRADATION **CALCINEURINPATHWAY** **COMPLEMENT AND COAGULATION CASCADES** **ECM RECEPTOR INTERACTION** TCRPATHWAY TNFR2PATHWAY **NO1PATHWAY** **INTRINSICPATHWAY** STATIN PATHWAY PHARMGKB ALKALOID BIOSYNTHESIS II LEUKOCYTE TRANSENDOTHELIAL MIGRATION HSA04514 CELL ADHESION MOLECULES INFLAMPATHWAY MEF2DPATHWAY IL2RBPATHWAY WNT BETA CATENIN PATHWAY HSA04510 FOCAL ADHESION NTHIPATHWAY **BCRPATHWAY** **BCR SIGNALING PATHWAY** **GCRPATHWAY** GLYCEROLIPID METABOLISM

	GPCRPATHWAY PDGFPATHWAY **INTEGRIN MEDIATED CELL ADHESION KEGG** FATTY ACID METABOLISM NKTPATHWAY CXCR4PATHWAY TOLLPATHWAY TNFR2PATHWAY LEUKOCYTE TRANSENDOTHELIAL MIGRATION PPAR SIGNALING PATHWAY AMYOTROPHIC LATERAL SCLEROSIS GLYCEROLIPID METABOLISM GLYCOSPHINGOLIPID METABOLISM MTORPATHWAY NICOTINATE AND NICOTINAMIDE METABOLISM PGC1APATHWAY METPATHWAY WNT BETA CATENIN PATHWAY TPOPATHWAY BUTANOATE METABOLISM TGF BETA SIGNALING PATHWAY GSK3PATHWAY GLYCOSAMINOGLYCAN DEGRADATION PROPANOATE METABOLISM CELL COMMUNICATION OVARIAN INFERTILITY GENES VALINE LEUCINE AND ISOLEUCINE DEGRADATION **BILE ACID BIOSYNTHESIS** IL6PATHWAY **BCR SIGNALING PATHWAY** CELL ADHESION MOLECULES B CELL RECEPTOR SIGNALING PATHWAY **BCRPATHWAY** PTENPATHWAY MCALPAINPATHWAY **1 AND 2 METHYLNAPHTHALENE DEGRADATION** GPCRDB CLASS B SECRETIN LIKE **IL2PATHWAY** PROPANOATE METABOLISM MELANOMA INTEGRINPATHWAY ALKALOID BIOSYNTHESIS II ETHER LIPID METABOLISM NUCLEAR RECEPTORS CK1PATHWAY TYPE I DIABETES MELLITUS **NO2IL12PATHWAY** **RHOPATHWAY** **CYTOKINE CYTOKINE RECEPTOR INTERACTION** BLOOD CLOTTING CASCADE INFLAMPATHWAY DICTYOSTELIUM DISCOIDEUM CAMP CHEMOTAXIS PATHWAY DIFFERENTIATION PATHWAY IN PC12 CELLS	**HISTIDINE METABOLISM** **GATA3PATHWAY** PEPTIDE GPCRS PIP3 SIGNALING IN B LYMPHOCYTES **CYTOKINE CYTOKINE RECEPTOR INTERACTION** INOSITOL PHOSPHATE METABOLISM **BADPATHWAY** MONOAMINE GPCRS NFKBPATHWAY ECMPATHWAY **1 AND 2 METHYLNAPHTHALENE DEGRADATION** B CELL RECEPTOR SIGNALING PATHWAY **VIPPATHWAY** **CARDIACEGFPATHWAY** **ALKPATHWAY** IL4RECEPTOR IN B LYPHOCYTES GHPATHWAY NATURAL KILLER CELL MEDIATED CYTOTOXICITY RAC1PATHWAY EDG1PATHWAY **BETA ALANINE METABOLISM** UCALPAINPATHWAY GPCRPATHWAY ERK1 ERK2 MAPK PATHWAY **INTEGRIN MEDIATED CELL ADHESION KEGG** ARACHIDONIC ACID METABOLISM ABC TRANSPORTERS GENERAL **RHOPATHWAY** JAK STAT SIGNALING PATHWAY HDACPATHWAY **BILE ACID BIOSYNTHESIS** TOLLPATHWAY SPPAPATHWAY **NDKDYNAMINPATHWAY** SMOOTH MUSCLE CONTRACTION CALCIUM SIGNALING PATHWAY GLYCEROLIPID METABOLISM ANTIGEN PROCESSING AND PRESENTATION ADIPOCYTOKINE SIGNALING PATHWAY
LAR	**CHOLESTEROL BIOSYNTHESIS** **PENTOSE AND GLUCURONATE INTERCONVERSIONS** **BIOSYNTHESIS OF STEROIDS** **TYROSINE METABOLISM** **GAMMA HEXACHLOROCYCLOHEXANE DEGRADATION** **PORPHYRIN AND CHLOROPHYLL METABOLISM** **PHENYLALANINE METABOLISM** **GAMMA HEXACHLOROCYCLOHEXANE DEGRADATION** **CHREBPPATHWAY** **GLUTATHIONE METABOLISM** 1 AND 2 METHYLNAPHTHALENE DEGRADATION GLYCOSPHINGOLIPID METABOLISM **PORPHYRIN AND CHLOROPHYLL METABOLISM**	**GLUTATHIONE METABOLISM** **PENTOSE AND GLUCURONATE INTERCONVERSIONS** **GLUTATHIONE METABOLISM** **CHOLESTEROL BIOSYNTHESIS** **TYROSINE METABOLISM** **BIOSYNTHESIS OF STEROIDS** **PORPHYRIN AND CHLOROPHYLL METABOLISM** **ANDROGEN AND ESTROGEN METABOLISM** **GLYCOSPHINGOLIPID METABOLISM** **TYPE III SECRETION SYSTEM** **GAMMA HEXACHLOROCYCLOHEXANE DEGRADATION** **FLAGELLAR ASSEMBLY** **CITRATE CYCLE TCA CYCLE** **PHENYLALANINE METABOLISM** **ATP SYNTHESIS** **PHOTOSYNTHESIS**

	ANDROGEN AND ESTROGEN METABOLISM **PHENYLALANINE METABOLISM** **CITRATE CYCLE TCA CYCLE** **TYROSINE METABOLISM** **METABOLISM OF XENOBIOTICS BY CYTOCHROME P450** **EICOSANOID SYNTHESIS** **VALINE LEUCINE AND ISOLEUCINE DEGRADATION** **GLYCAN STRUCTURES DEGRADATION** GLYCOSYLPHOSPHATIDYLINOSITOL ANCHOR BIOSYNTHESIS BUTANOATE METABOLISM **ARGININE AND PROLINE METABOLISM** **GLUTATHIONE METABOLISM** **CK1PATHWAY** **CITRATE CYCLE** **PANTOTHENATE AND COA BIOSYNTHESIS** PROPANOATE METABOLISM **FATTY ACID METABOLISM** **ANDROGEN AND ESTROGEN METABOLISM** **GLUTAMATE METABOLISM** **N GLYCAN BIOSYNTHESIS** **HISTIDINE METABOLISM** BILE ACID BIOSYNTHESIS PROPANOATE METABOLISM HCMVPATHWAY **TRYPTOPHAN METABOLISM** **ALANINE AND ASPARTATE METABOLISM** VALINE LEUCINE AND **ISOLEUCINE** DEGRADATION **PPAR SIGNALING PATHWAY** **CREBPATHWAY** **ATP SYNTHESIS** **FLAGELLAR ASSEMBLY** **TYPE III SECRETION SYSTEM** SPRYPATHWAY ERYTHPATHWAY **PHOTOSYNTHESIS** **HISTIDINE METABOLISM** MCALPAINPATHWAY **FRUCTOSE AND MANNOSE METABOLISM** GLYCOSAMINOGLYCAN DEGRADATION **SPHINGOLIPID METABOLISM** **GLUTAMATE METABOLISM** **LINOLEIC ACID METABOLISM** GLYCEROLIPID METABOLISM **ARGININE AND PROLINE METABOLISM** SNARE INTERACTIONS IN VESICULAR TRANSPORT SELENOAMINO ACID METABOLISM **PENTOSE PHOSPHATE PATHWAY** **FRUCTOSE AND MANNOSE METABOLISM** NO1PATHWAY BILE ACID BIOSYNTHESIS **STARCH AND SUCROSE METABOLISM** CYSTEINE METABOLISM GATA3PATHWAY **ALANINE AND ASPARTATE METABOLISM** LIMONENE AND PINENE DEGRADATION ABC TRANSPORTERS GENERAL UREA CYCLE AND METABOLISM OF AMINO GROUPS **OXIDATIVE PHOSPHORYLATION** **N GLYCAN BIOSYNTHESIS** GLYCEROLIPID METABOLISM **KREBS TCA CYCLE** AMINOSUGARS METABOLISM **ARACHIDONIC ACID METABOLISM** **TRYPTOPHAN METABOLISM** **PENTOSE PHOSPHATE PATHWAY** NAPHTHALENE AND ANTHRACENE DEGRADATION **OXIDATIVE PHOSPHORYLATION**	**STARCH AND SUCROSE METABOLISM** **PORPHYRIN AND CHLOROPHYLL METABOLISM** **ARGININE AND PROLINE METABOLISM** **GAMMA HEXACHLOROCYCLOHEXANE DEGRADATION** **METABOLISM OF XENOBIOTICS BY CYTOCHROME P450** **FRUCTOSE AND MANNOSE METABOLISM** CARBON FIXATION **GLYCAN STRUCTURES DEGRADATION** **CITRATE CYCLE** COMPPATHWAY PARKINSONS DISEASE **ANDROGEN AND ESTROGEN METABOLISM** **PANTOTHENATE AND COA BIOSYNTHESIS** **FATTY ACID METABOLISM** RENIN ANGIOTENSIN SYSTEM **TYROSINE METABOLISM** **GLUTAMATE METABOLISM** **ALANINE AND ASPARTATE METABOLISM** CARBON FIXATION **EICOSANOID SYNTHESIS** GLYCOSAMINOGLYCAN DEGRADATION **PHENYLALANINE METABOLISM** **CHREBPPATHWAY** **CK1PATHWAY** MONOAMINE GPCRS **SPHINGOLIPID METABOLISM** P53HYPOXIAPATHWAY **TRYPTOPHAN METABOLISM** NICOTINATE AND NICOTINAMIDE METABOLISM **GLUTAMATE METABOLISM** UCALPAINPATHWAY **FRUCTOSE AND MANNOSE METABOLISM** **KREBS TCA CYCLE** **OXIDATIVE PHOSPHORYLATION** **HISTIDINE METABOLISM** GALACTOSE METABOLISM **PENTOSE PHOSPHATE PATHWAY** **OXIDATIVE PHOSPHORYLATION** **N GLYCAN BIOSYNTHESIS** **LINOLEIC ACID METABOLISM** BLOOD CLOTTING CASCADE **ARACHIDONIC ACID METABOLISM** MCALPAINPATHWAY ERYTHPATHWAY ABC TRANSPORTERS GENERAL BUTANOATE METABOLISM O GLYCAN BIOSYNTHESIS **PPAR SIGNALING PATHWAY** **ALANINE AND ASPARTATE METABOLISM** **VALINE LEUCINE AND ISOLEUCINE DEGRADATION** NO1PATHWAY **STARCH AND SUCROSE METABOLISM** **ARGININE AND PROLINE METABOLISM** COMPLEMENT AND COAGULATION CASCADES GALACTOSE METABOLISM **PENTOSE PHOSPHATE PATHWAY**

UNC	DNA REPLICATION REACTOME ATRBRCAPATHWAY DNA POLYMERASE CELL CYCLE KEGG CELLCYCLEPATHWAY G1 TO S CELL CYCLE REACTOME CELL CYCLE G1PATHWAY G1 AND S PHASES BASAL TRANSCRIPTION FACTORS RNA TRANSCRIPTION REACTOME AMINOACYL TRNA BIOSYNTHESIS RNA POLYMERASE PYRIMIDINE METABOLISM CARM ERPATHWAY PYRIMIDINE METABOLISM ALANINE AND ASPARTATE METABOLISM ALANINE AND ASPARTATE METABOLISM AMINOACYL TRNA BIOSYNTHESIS G2PATHWAY P53PATHWAY SELENOAMINO ACID METABOLISM GLUTAMATE METABOLISM PENTOSE PHOSPHATE PATHWAY MRNA PROCESSING REACTOME PENTOSE PHOSPHATE PATHWAY	

Bold- shared between training and validation set