Genomic Analyses across Six Cancer Types Identify Basal-like Breast Cancer as a Unique Molecular Entity

Aleix Prat, Barbara Adamo, Cheng Fan, Vicente Peg, Maria Vidal, Patricia Galván, Ana Vivancos, Paolo Nuciforo, Héctor G. Palmer, Shaheenah Dawood, Jordi Rodón, Santiago Ramon y Cajal, Josep Maria Del Campo, Enriqueta Felip, Josep Tabernero and Javier Cortés

Supplemental Information

Fig. S1 Principal component loading plots for PC1, PC2 and PC3 in an independent microarray dataset (GSE23768) of breast cancers, colorectal adenocarcinomas, lung adenocarcinoma and lung squamous cell lung cancer. Samples are colored based on cancer type.

Fig. S2 Consensus clustering analysis. (**A**) Cumulative distribution function (CDF) for k=2 to k=10. (**B**) Delta area plot.

Fig. S3 Ability of a previously published TP53 mutation vs. wild-type signature to predict TP53 mutations in the TCGA combined dataset. (**A**) TP53 signature scores across TP53WT vs. TP53 mutated cancers. (**B**) Area under the ROC curve of the signature for predicting TP53 mutations. (**C**) Distribution of TP53 mutation percentages across cancer types (with gene expression data) and area under the ROC curve in each category.

Fig. S4 Correlation of a previously published PTEN-loss gene signature with PTEN, INPP4B , phosphor-S6 and phosphor-4E-BP1 protein expression in the TCGA breast cancer dataset.

Fig. S5 PAM50 intrinsic subtyping of an independent microarray dataset (GSE23768) of breast cancers, colorectal adenocarcinomas, lung adenocarcinoma and lung squamous cell lung cancer. Below the array tree, subtype calls are provided, including Claudin-low vs. not calls.

Fig. S6 PAM50 intrinsic subtyping of an in-house breast cancer and colorectal adenocarcinoma dataset. Below the array tree, subtype calls are provided. Note: only 49 genes of the 50 PAM50 genes were evaluated.

Fig. S7 HER2 FISH of a CRC sample from Fig. S6.

Fig. S8 Evaluation of batch effects across 4 TCGA cancer types during year 2000. (**A**) Distribution of microarray chip barcodes across 4 cancer types. No significant differences were observed. LUAD, lung adenocarcinoma; SCC, squamous cell lung cancer. (**B**) Principal component loading plots for PC1, PC2 and PC3. Samples are colored based on cancer type. (**C**) Principal component loading plots for PC1, PC2 and PC3. Samples are colored based on the rank ordered microarray chip barcodes divided into deciles.

Fig. S9 Evaluation of batch effects within each TCGA cancer type. Principal component loading plots for PC1 and PC2. Samples are colored based on the rank ordered microarray chip barcodes divided into two groups (i.e. above and below the median of all barcode numbers).

Figure S1

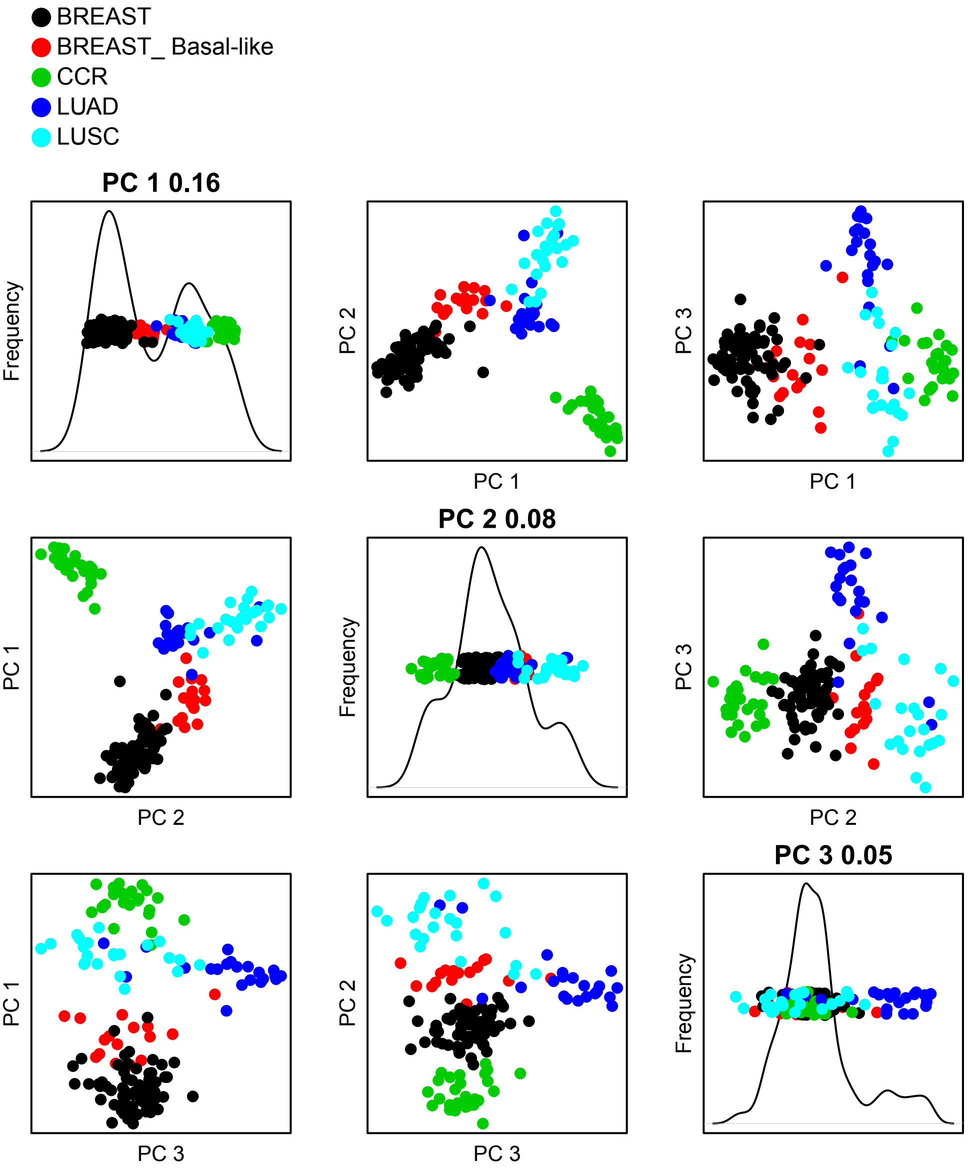

● BREAST
● BREAST_ Basal-like
● CCR
● LUAD
● LUSC
PC 1 0.16
Frequency
PC 2
PC 1
PC 3
PC 1
PC 1
PC 2
PC 2 0.08
Frequency
PC 3
PC 2
PC 1
PC 3
PC 2
PC 3
PC 3 0.05
Frequency

Figure S2

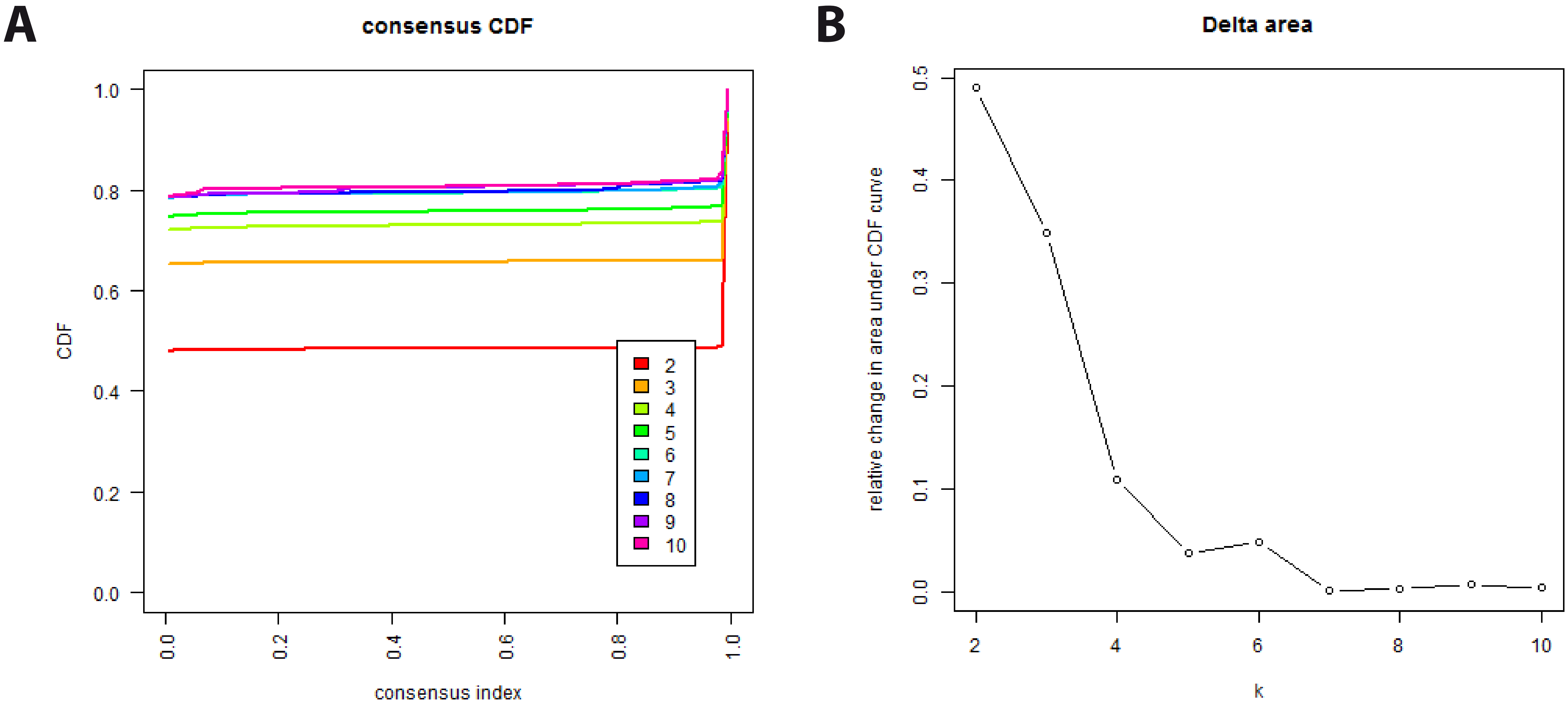
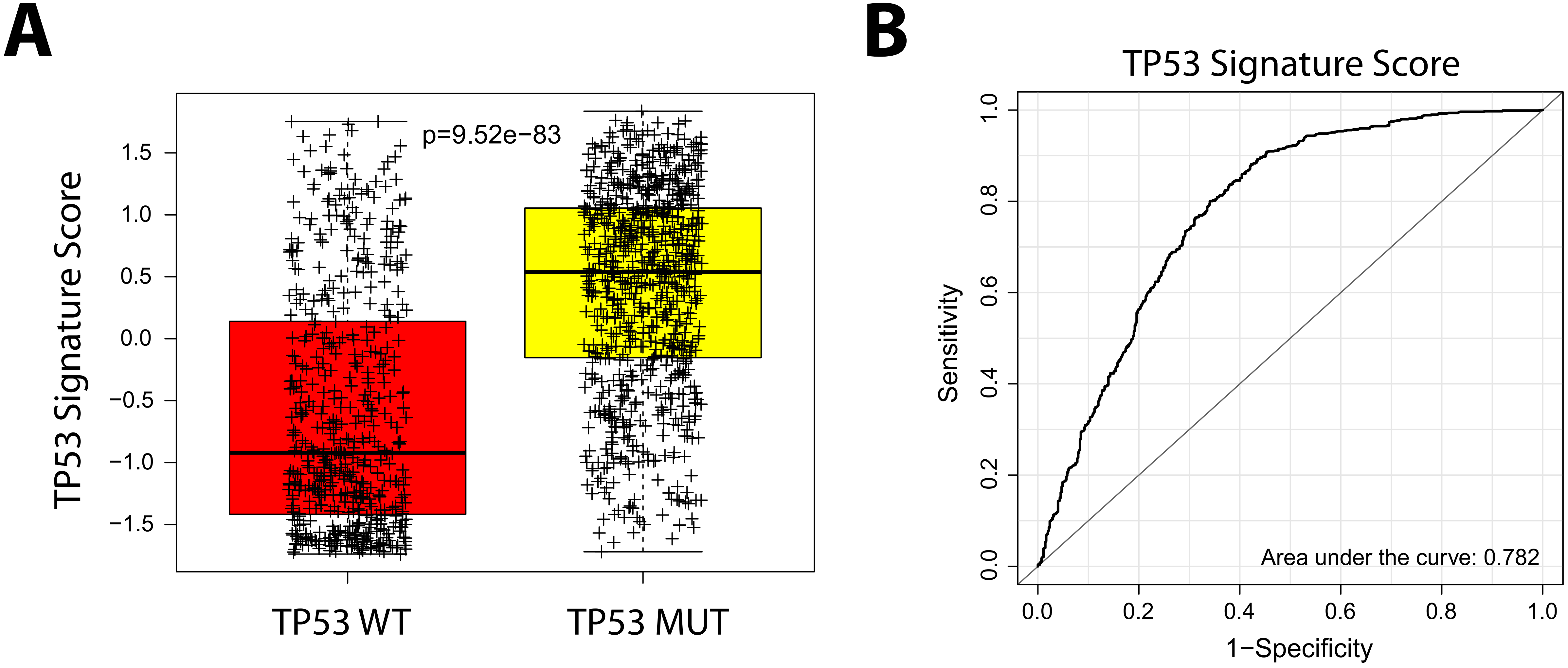

A
consensus CDF
CDF
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
consensus index
2
3
4
5
6
7
8
9
10
B
Delta area
relative change in area under CDF curve
0.5
0.4
0.3
0.2
0.1
0.0
2
4
6
8
10
k

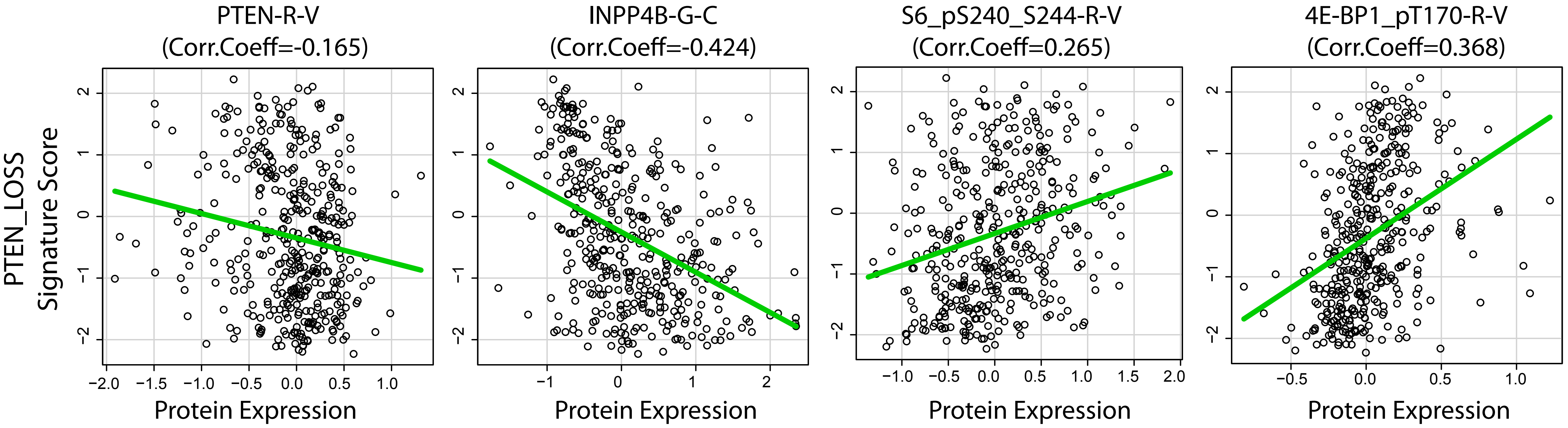
Figure S3



C

	TP53MUT	%	TP53WT	%	TOTAL	TP53 Signature Score Area under the ROC curve
All samples	797	59%	552	41%	1349	0.782
OVARIAN	303	96%	13	4%	316	0.71
BREAST (NON-BASAL-LIKE)	113	27%	303	73%	416	0.802
BREAST BASAL-LIKE	76	82%	17	18%	93	0.587
BREAST (ALL)	189	37%	320	63%	509	0.847
CCR	112	54%	94	46%	206	0.71
GBM	30	40%	45	60%	75	0.649
LUAD	62	51%	59	49%	121	0.748
SQCLC	101	83%	21	17%	122	0.609
LUNG (ALL)	92	47%	104	53%	196	0.749

Figure S4



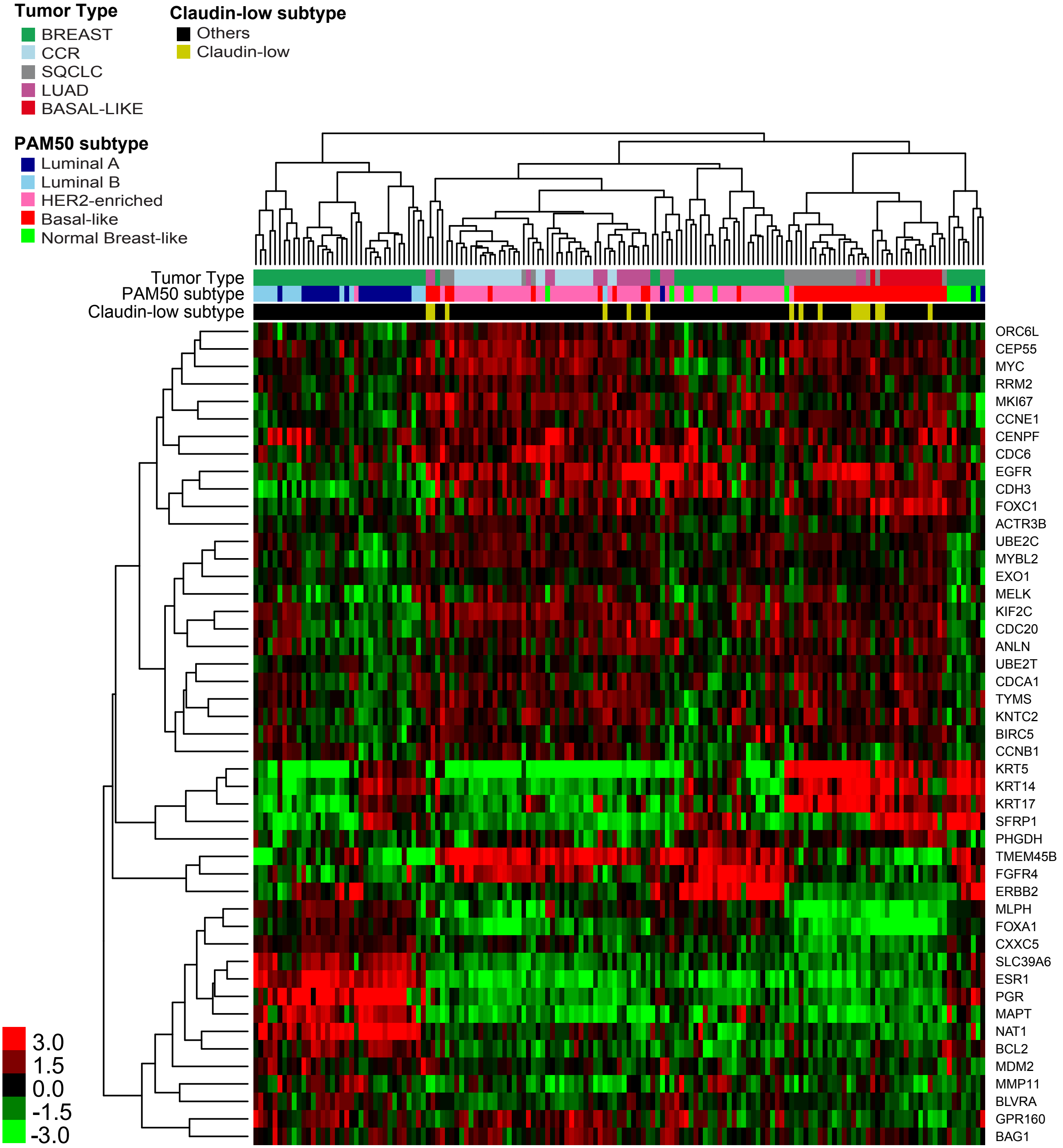
PTEN-R-V (Corr.Coeff=-0.165)

INPP4B-G-C (Corr.Coeff=-0.424)

S6_pS240_S244-R-V (Corr.Coeff=0.265)

4E-BP1_pT170-R-V (Corr.Coeff=0.368)

Figure S5



Tumor Type
BREAST
CCR
SQCLC
LUAD
BASAL-LIKE
Claudin-low subtype
Others
Claudin-low
PAM50 subtype
Luminal A
Luminal B
HER2-enriched
Basal-like
Normal Breast-like
Tumor Type
PAM50 subtype
Claudin-low subtype
ORC6L
CEP55
MYC
RRM2
MKI67
CCNE1
CENPF
CDC6
EGFR
CDH3
FOXC1
ACTR3B
UBE2C
MYBL2
EXO1
MELK
KIF2C
CDC20
ANLN
UBE2T
CDCA1
TYMS
KNTC2
BIRC5
CCNB1
KRT5
KRT14
KRT17
SFRP1
PHGDH
TMEM45B
FGFR4
ERBB2
MLPH
FOXA1
CXXC5
SLC39A6
ESR1
PGR
MAPT
NAT1
BCL2
MDM2
MMP11
BLVRA
GPR160
BAG1
3.0
1.5
0.0
-1.5
-3.0

Figure S6

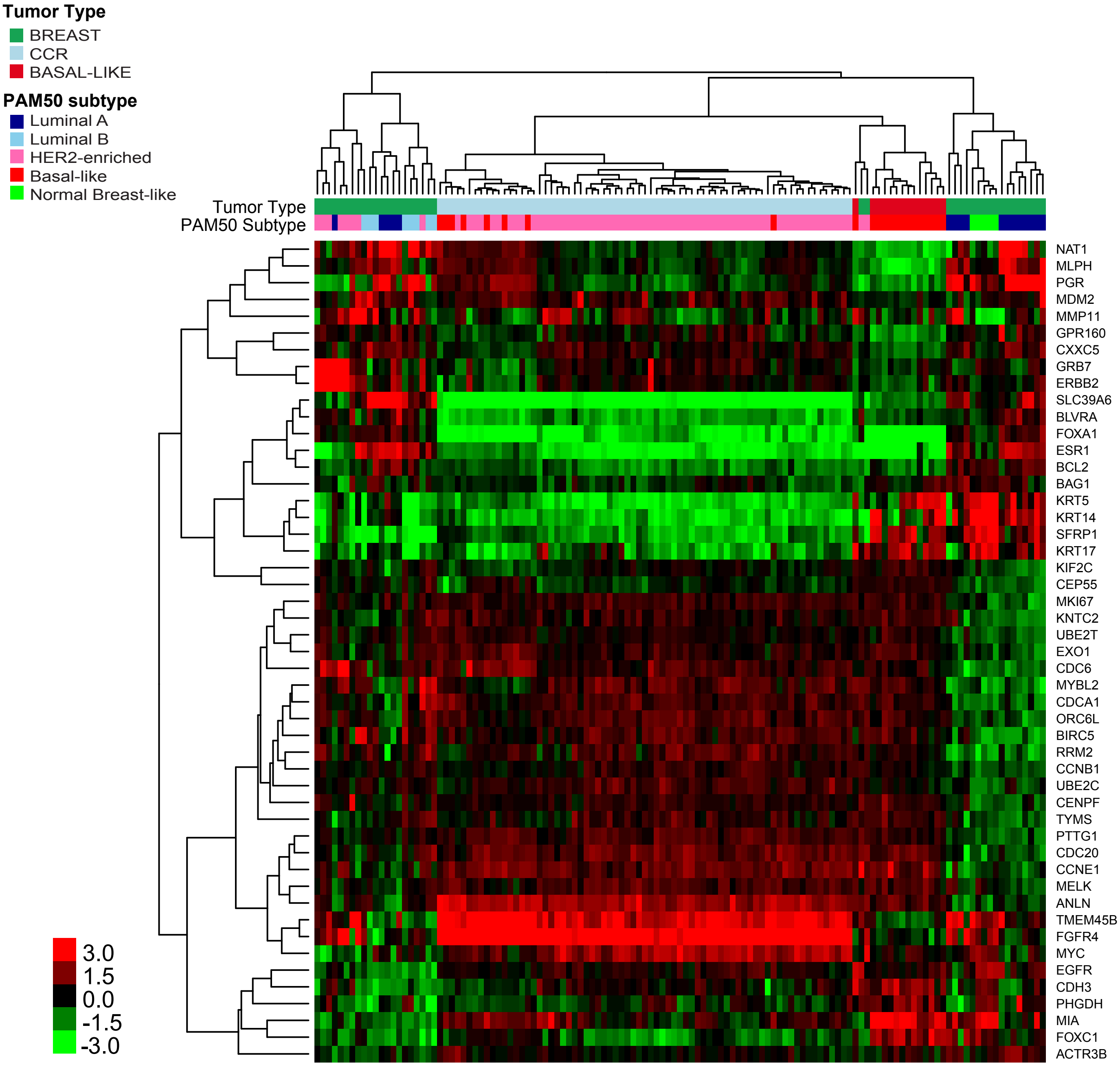

Tumor Type
BREAST
CCR
BASAL-LIKE
PAM50 subtype
Luminal A
Luminal B
HER2-enriched
Basal-like
Normal Breast-like
Tumor Type
PAM50 Subtype
NAT1
MLPH
PGR
MDM2
MMP11
GPR160
CXXC5
GRB7
ERBB2
SLC39A6
BLVRA
FOXA1
ESR1
BCL2
BAG1
KRT5
KRT14
SFRP1
KRT17
KIF2C
CEP55
MKI67
KNTC2
UBE2T
EXO1
CDC6
MYBL2
CDCA1
ORC6L
BIRC5
RRM2
CCNB1
UBE2C
CENPF
TYMS
PTTG1
CDC20
CCNE1
MELK
ANLN
TMEM45B
FGFR4
MYC
EGFR
CDH3
PHGDH
MIA
FOXC1
ACTR3B
3.0
1.5
0.0
-1.5
-3.0

Figure S7

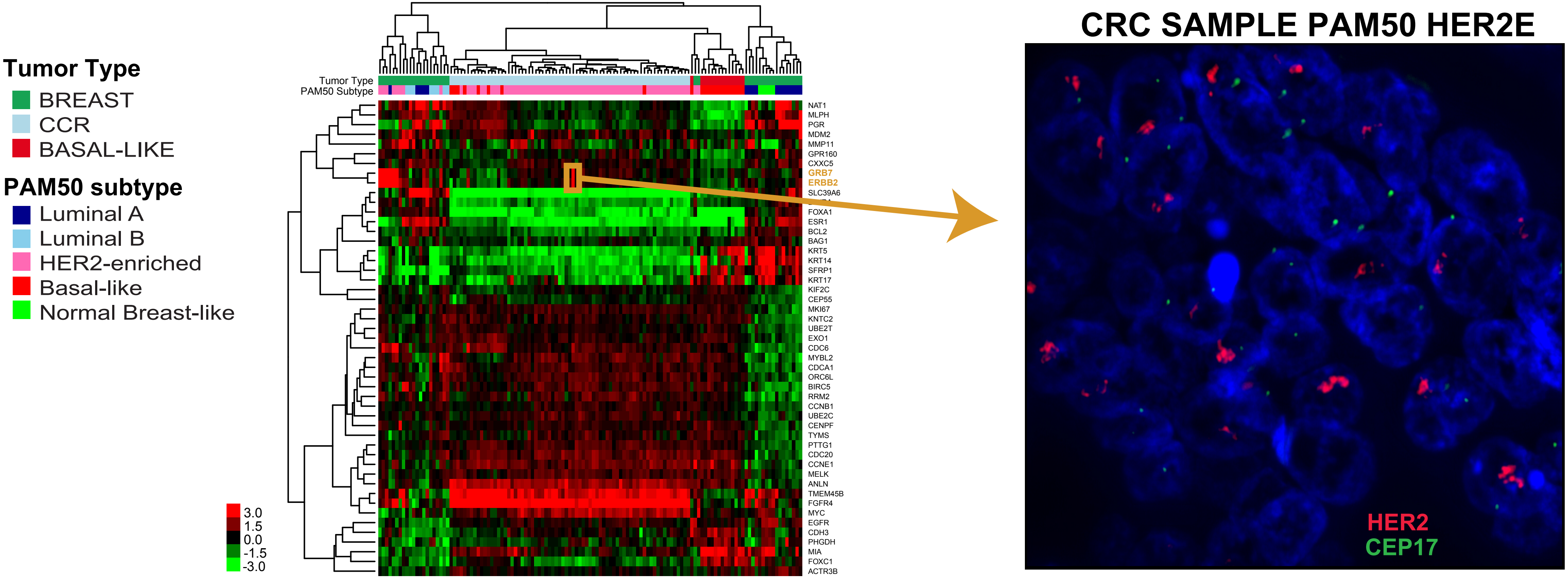

Tumor Type
🟩 BREAST
🟦 CCR
🟥 BASAL-LIKE
PAM50 subtype
🟦 Luminal A
🟦 Luminal B
🟪 HER2-enriched
🟥 Basal-like
🟩 Normal Breast-like
Tumor Type
PAM50 Subtype
NAT1
MLPH
PGR
MDM2
MMP11
GPR160
CXXC5
GRB7
ERBB2
SLC39A6
FOXA1
ESR1
BCL2
BAG1
KRT5
KRT14
SFRP1
KRT17
KIF2C
CEP55
MKI67
KNTC2
UBE2T
EXO1
CDC6
MYBL2
CDCA1
ORC6L
BIRC5
RRM2
CCNB1
UBE2C
CENPF
TYMS
PTTG1
CDC20
CCNE1
MELK
ANLN
TMEM45B
FGFR4
MYC
EGFR
CDH3
PHGDH
MIA
FOXC1
ACTR3B
3.0
1.5
0.0
-1.5
-3.0
CRC SAMPLE PAM50 HER2E
HER2
CEP17

Figure S8

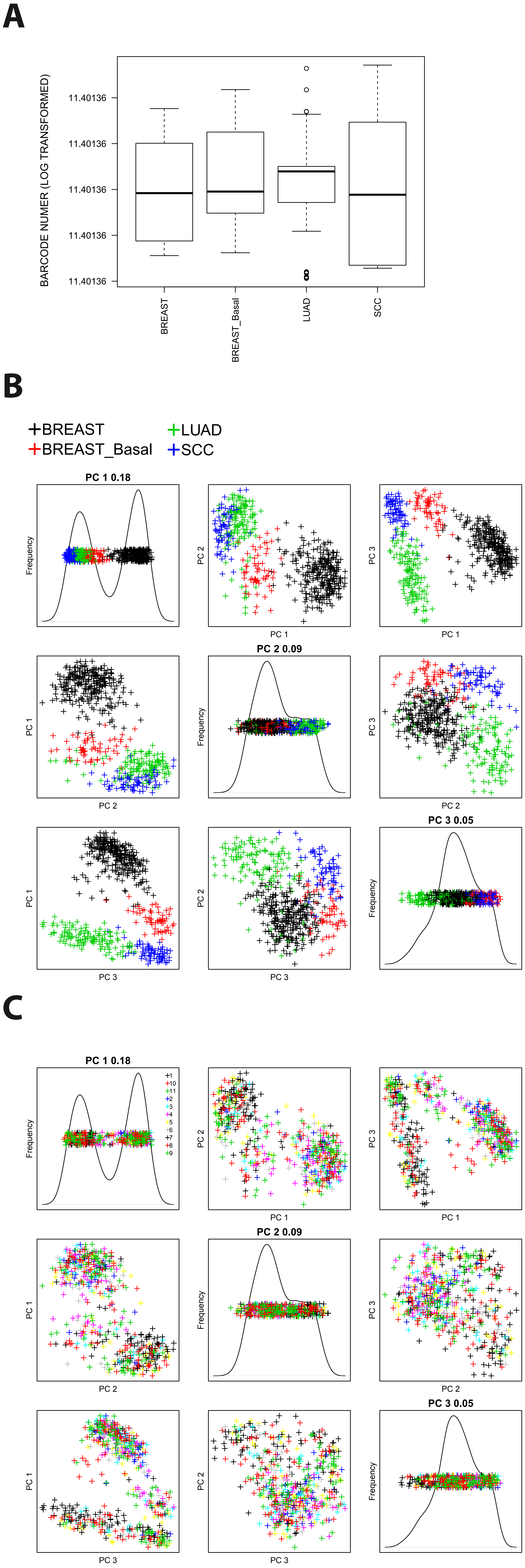
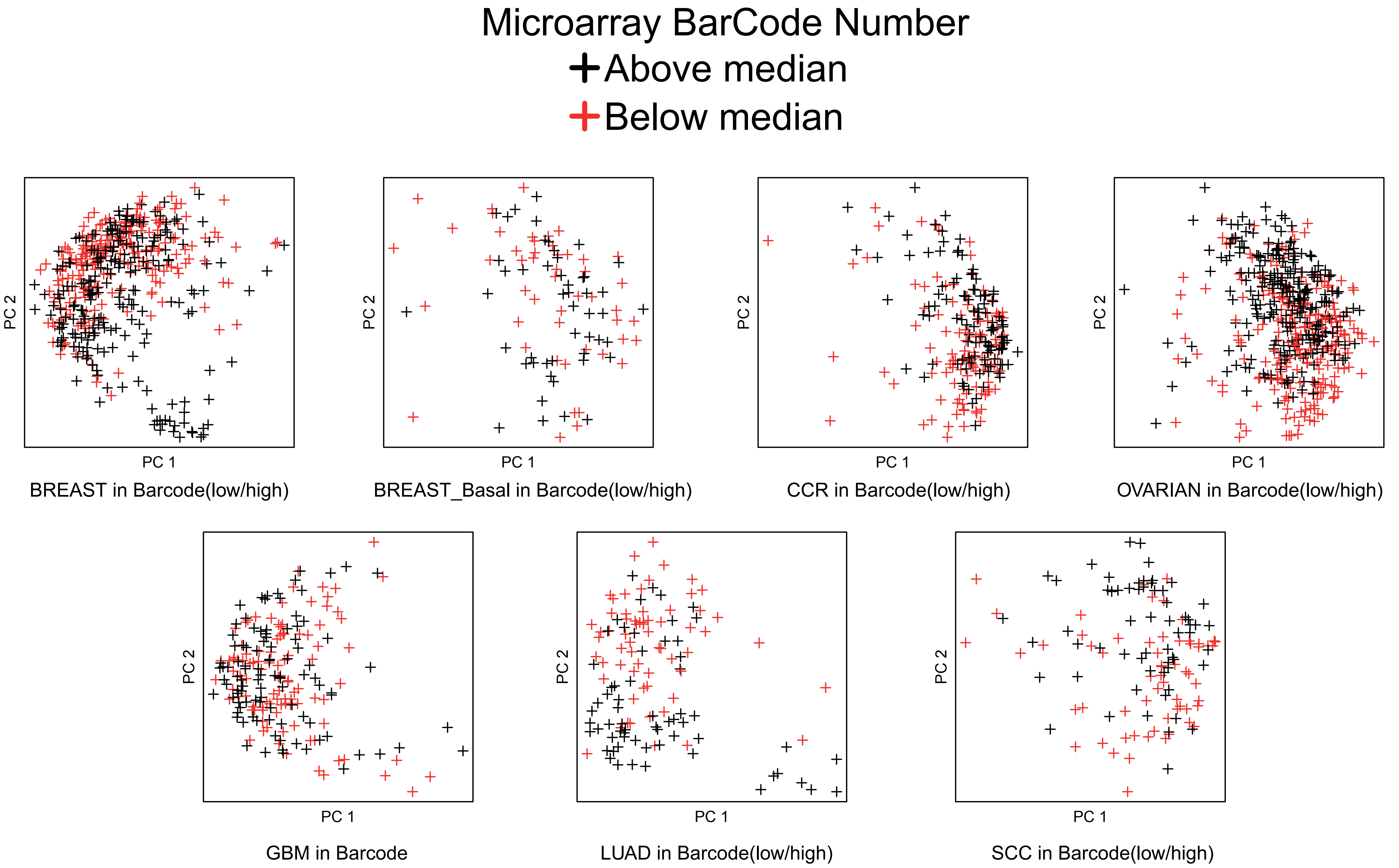

A
BARCODE NUMER (LOG TRANSFORMED)
11.40136
11.40136
11.40136
11.40136
11.40136
BREAST
BREAST_Basal
LUAD
SCC
B
+BREAST
+BREAST_Basal
+LUAD
+SCC
PC 1 0.18
Frequency
PC 2
PC 1
PC 3
PC 1
PC 1
PC 2
PC 2 0.09
Frequency
PC 3
PC 2
PC 1
PC 3
PC 2
PC 3
PC 3 0.05
Frequency
C
PC 1 0.18
+1
+10
+11
+2
+3
+4
+5
+6
+7
+8
+9
Frequency
PC 2
PC 1
PC 3
PC 1
PC 1
PC 2
PC 2 0.09
Frequency
PC 3
PC 2
PC 1
PC 3
PC 2
PC 3
PC 3 0.05
Frequency

Figure S9



BREAST in Barcode(low/high)

BREAST_Basal in Barcode(low/high)

CCR in Barcode(low/high)

OVARIAN in Barcode(low/high)

GBM in Barcode

LUAD in Barcode(low/high)

SCC in Barcode(low/high)