Cell Reports, Volume *42*

Supplemental information

Heterogeneity and transcriptional drivers of triple-negative breast cancer

Bojana Jovanović, Daniel Temko, Laura E. Stevens, Marco Seehawer, Anne Fassl, Katherine Murphy, Jayati Anand, Kodie Garza, Anushree Gulvady, Xintao Qiu, Nicholas W. Harper, Veerle W. Daniels, Huang Xiao-Yun, Jennifer Y. Ge, Maša Alečković, Jason Pyrdol, Kunihiko Hinohara, Shawn B. Egri, Malvina Papanastasiou, Raga Vadhi, Alba Font-Tello, Robert Witwicki, Guillermo Peluffo, Anne Trinh, Shaokun Shu, Benedetto Diciaccio, Muhammad B. Ekram, Ashim Subedee, Zachary T. Herbert, Kai W. Wucherpfennig, Anthony G. Letai, Jacob D. Jaffe, Piotr Sicinski, Myles Brown, Deborah Dillon, Henry W. Long, Franziska Michor, and Kornelia Polyak

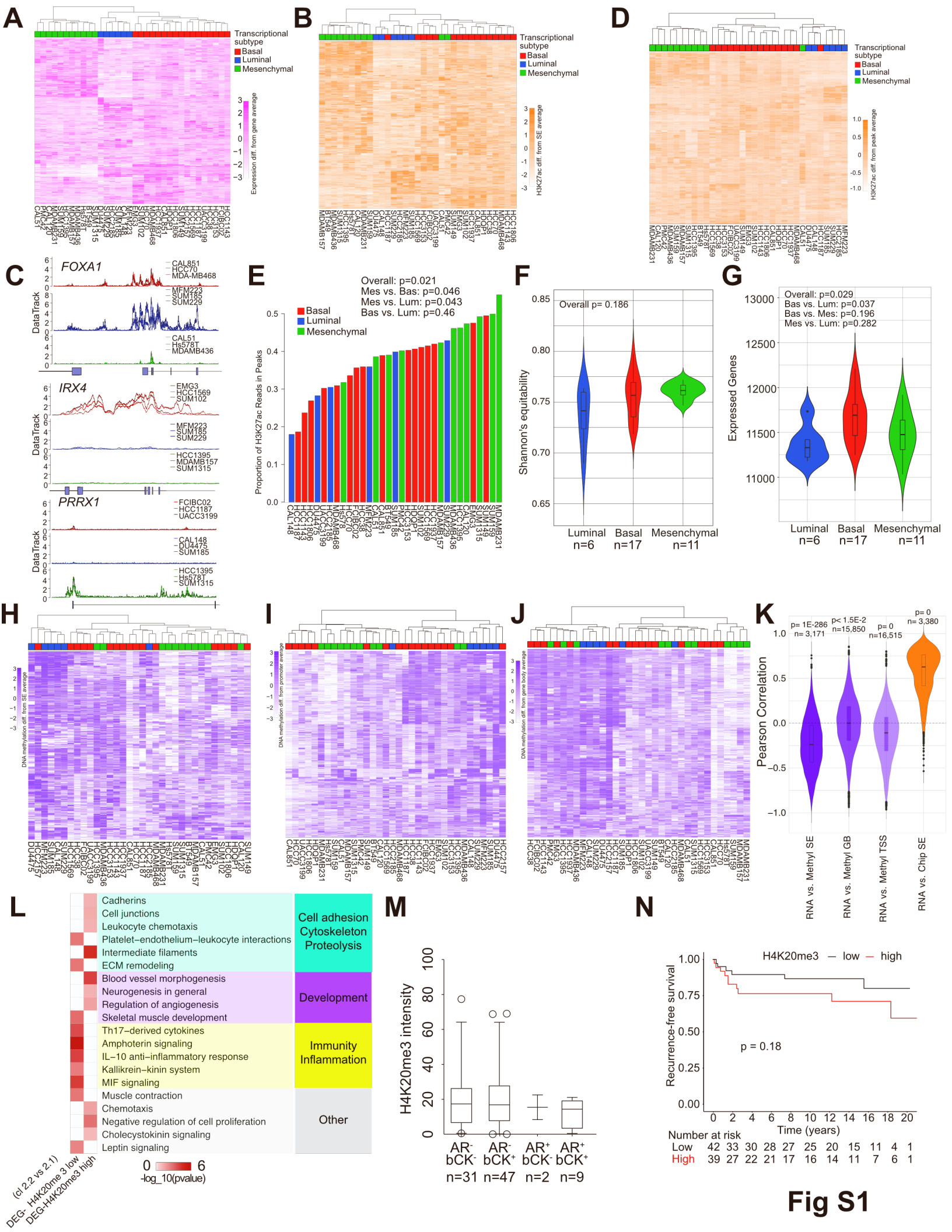
A
Transcriptional subtype
Basal
Luminal
Mesenchymal
Expression diff. from gene average
B
Transcriptional subtype
Basal
Luminal
Mesenchymal
H3K27ac diff. from SE average
D
Transcriptional subtype
Basal
Luminal
Mesenchymal
H3K27ac diff. from peak average
C
FOXA1
CAL851
HCC70
MDA-MB468
MFM223
SUM185
SUM229
CAL51
Hs578T
MDAMB436
IRX4
EMG3
HCC1569
SUM102
MFM223
SUM185
SUM229
HCC1395
MDAMB157
SUM1315
PRRX1
FCIBC02
HCC1187
UACC3199
CAL148
DU4475
SUM185
HCC1395
Hs578T
SUM1315
DataTrack
E
Overall: p=0.021
Mes vs. Bas: p=0.046
Mes vs. Lum: p=0.043
Bas vs. Lum: p=0.46
Basal
Luminal
Mesenchymal
Proportion of H3K27ac Reads in Peaks
F
Overall p= 0.186
Shannon's equitability
Luminal n=6
Basal n=17
Mesenchymal n=11
G
Overall: p=0.029
Bas vs. Lum: p=0.037
Bas vs. Mes: p=0.196
Mes vs. Lum: p=0.282
Expressed Genes
Luminal n=6
Basal n=17
Mesenchymal n=11
H
DNA methylation diff. from SE average
I
DNA methylation diff. from promoter average
J
DNA methylation diff. from gene body average
K
p= 1E-286
n= 3,171
p< 1.5E-2
n=15,850
p= 0
n=16,515
p= 0
n= 3,380
Pearson Correlation
RNA vs. Methyl SE
RNA vs. Methyl GB
RNA vs. Methyl TSS
RNA vs. Chip SE
L
Cadherins
Cell junctions
Leukocyte chemotaxis
Platelet–endothelium–leukocyte interactions
Intermediate filaments
ECM remodeling
Blood vessel morphogenesis
Neurogenesis in general
Regulation of angiogenesis
Skeletal muscle development
Th17–derived cytokines
Amphoterin signaling
IL–10 anti–inflammatory response
Kallikrein–kinin system
MIF signaling
Muscle contraction
Chemotaxis
Negative regulation of cell proliferation
Cholecystokinin signaling
Leptin signaling
Cell adhesion Cytoskeleton Proteolysis
Development
Immunity Inflammation
Other
(cl 2,2 vs 2,1)
DEG- H4K20me 3 low
DEG-H4K20me3 high
-log_10(pvalue)
M
H4K20me3 intensity
AR- bCK- n=31
AR- bCK+ n=47
AR+ bCK- n=2
AR+ bCK+ n=9
N
H4K20me3 — low — high
p = 0.18
Recurrence-free survival
Time (years)
Number at risk
Low 42 33 30 28 27 25 20 15 11 4 1
High 39 27 22 21 17 16 14 11 7 6 1

Fig S1

Figure S1. TNBC transcriptional and epigenetic subtypes. Related to Figure 1.

(A) Heatmap showing clustering of 34 TNBC cell lines based on the expression of the top 20% most variable genes. Values shown are expression differences from gene average. Values are capped at +/- 3 for the purpose of visualization.

(B) Heatmap showing clustering of 33 TNBC cell lines based on H3K27ac signal in the top 20% most variable super-enhancers. Values are capped at +/-3 for the purpose of visualization.

(C) Examples of subtype-specific super-enhancer regions.

(D) Heatmap showing clustering of 33 TNBC cell lines based on H3K27ac signal in the top 10% most variable peaks.

(E) Proportion of H3K27ac reads in peaks in individual TNBC cell lines. Overall p value (p= 0.021) from Kruskal-Wallis test, pairwise p values (p=0.046 mes vs bas, p=0.043 mes vs lum, p=0.46 bas vs lum) from Dunn's test, adjusted using Holm's method.

(F) Violin plot of Shannon's equitability across cell lines of each TNBC type. Bottom and top hinges of inset box plots show the 25th and 75th percentiles. Upper whisker extends from the upper hinge to the highest value that is no further than 1.5 times the interquartile range (IQR) from the hinge. Lower whisker extends from the lower hinge to the lowest value no further than 1.5 times the IQR from the hinge. Any data beyond the end of the whiskers are plotted as individual points. Overall p value (p=0.186) from Kruskal-Wallis test.

(G) Violin plot of number of expressed genes (FPKM > 1) across cell lines of each TNBC type. Bottom and top hinges of inset box plots show the 25th and 75th percentiles. Upper whisker extends from the upper hinge to the highest value that is no further than 1.5 times the interquartile range (IQR) from the hinge. Lower whisker extends from the lower hinge to the lowest value no further than 1.5 times the IQR from the hinge. Any data beyond the end of the whiskers are plotted as individual points. Overall p value (p=0.029) from Kruskal-Wallis test, pairwise p values (p=0.0.037 bas vs lum, p=0.196 bas vs mes, p=0.282 mes vs lum,) from Dunn's test, adjusted using Holm's method.

(H) Heatmap showing clustering of 34 TNBC cell lines based on DNA methylation levels in the top 20% most variable super-enhancers. Values shown are methylation differences from super-enhancer average. Values are capped at +/-3 for the purpose of visualization.

(I) Heatmap showing clustering of 34 TNBC cell lines based on DNA methylation levels in top 20% most variable promoters. Values shown are methylation differences from promoter average. Values are capped at +/- 3 for the purposes of visualization.

(J) Heatmap showing clustering of 34 TNBC cell lines based on DNA methylation levels in the top 20% most variable gene bodies. Values shown are methylation differences from gene body average. Values are capped at +/- 3 for the purposes of visualization.

(K) Violin plot of Pearson correlations between gene expression level, and DNA methylation and H3K27ac levels in regulatory elements for individual genes. GB: Gene body, SE: Super-enhancer, TSS: Gene promoter. Mann-Whitney U test p values are shown. DNA methylation in SEs median correlation across genes = -0.239. H3K27ac in SEs median correlation across genes = 0.626.

(L) Heatmap depicting Metacore network enrichment in differentially expressed genes (DEG) between cell lines in histone mass spectrometry cluster 2.2 versus cell lines in histone mass spectrometry cluster 2.1 from Figure 1I. DGE low/high refers to H4K20me3 levels in the two clusters.

(M) Box plot demonstrating H4K20me3 signal intensity across TNBC patient tumors positive or negative for AR and/or bCK. No significant differences were observed among the groups (ANOVA test p=0.5111). Bottom and top hinges of inset box plots show the 25th and 75th percentiles. Upper whisker extends from the upper hinge to the highest value that is no further than 1.5 times the interquartile range (IQR) from the hinge. Lower whisker extends from the lower hinge to the lowest value no further than 1.5 times the IQR from the hinge. Any data beyond the end of the whiskers are plotted as individual points.

(N) Kaplan-Meier plot depicting differences in recurrence-free survival (RSF) of TNBC patients with low or high H4K20me3 levels based on median-split of H4K20me3 immunofluorescence-based intensity. Number of patients at risk over time are depicted under graph (p=0.18, Cox regression).

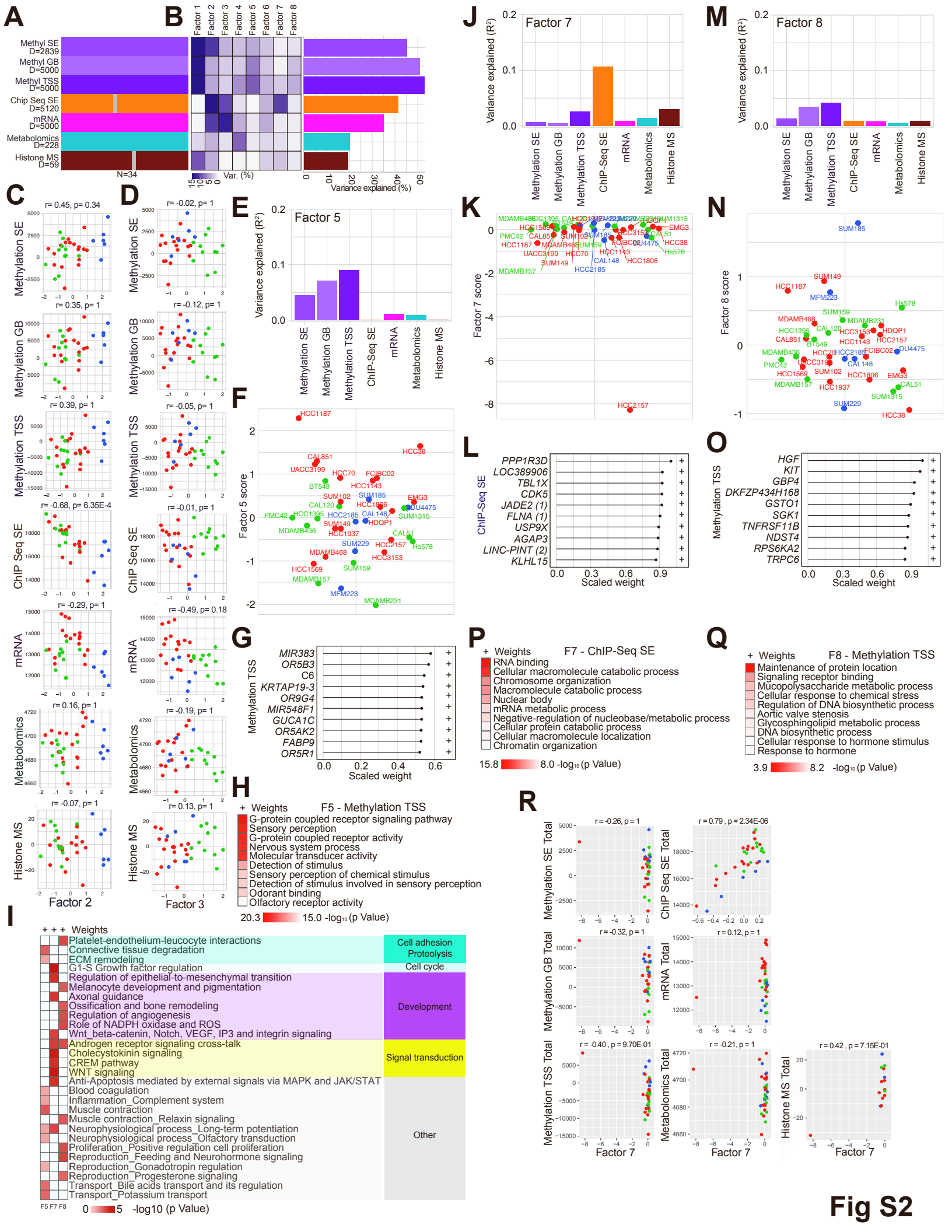
A
Methyl SE D=2839
Methyl GB D=5000
Methyl TSS D=5000
Chip Seq SE D=5120
mRNA D=5000
Metabolomics D=228
Histone MS D=59
N=34
B
Factor 1
Factor 2
Factor 3
Factor 4
Factor 5
Factor 6
Factor 7
Factor 8
Var. (%)
Variance explained (%)
C
r= 0.45, p= 0.34
r= 0.35, p= 1
r= 0.39, p= 1
r= -0.68, p= 6.35E-4
r= -0.29, p= 1
r= 0.16, p= 1
r= -0.07, p= 1
Methylation SE
Methylation GB
Methylation TSS
ChIP Seq SE
mRNA
Metabolomics
Histone MS
Factor 2
D
r= -0.02, p= 1
r= -0.12, p= 1
r= -0.05, p= 1
r= -0.01, p= 1
r= -0.49, p= 0.18
r= -0.19, p= 1
r= 0.13, p= 1
Factor 3
E
Factor 5
Variance explained (R²)
F
Factor 5 score
G
Methylation TSS
MIR383
OR5B3
C6
KRTAP19-3
OR9G4
MIR548F1
GUCA1C
OR5AK2
FABP9
OR5R1
Scaled weight
H
+ Weights
F5 - Methylation TSS
G-protein coupled receptor signaling pathway
Sensory perception
G-protein coupled receptor activity
Nervous system process
Molecular transducer activity
Detection of stimulus
Sensory perception of chemical stimulus
Detection of stimulus involved in sensory perception
Odorant binding
Olfactory receptor activity
20.3
15.0 -log10 (p Value)
I
+ + + Weights
Platelet-endothelium-leucocyte interactions
Connective tissue degradation
ECM remodeling
G1-S Growth factor regulation
Regulation of epithelial-to-mesenchymal transition
Melanocyte development and pigmentation
Axonal guidance
Ossification and bone remodeling
Regulation of angiogenesis
Role of NADPH oxidase and ROS
Wnt_beta-catenin, Notch, VEGF, IP3 and integrin signaling
Androgen receptor signaling cross-talk
Cholecystokinin signaling
CREM pathway
WNT signaling
Anti-Apoptosis mediated by external signals via MAPK and JAK/STAT
Blood coagulation
Inflammation_Complement system
Muscle contraction
Muscle contraction_Relaxin signaling
Neurophysiological process_Long-term potentiation
Neurophysiological process_Olfactory transduction
Proliferation_Positive regulation cell proliferation
Reproduction_Feeding and Neurohormone signaling
Reproduction_Gonadotropin regulation
Reproduction_Progesterone signaling
Transport_Bile acids transport and its regulation
Transport_Potassium transport
Cell adhesion
Proteolysis
Cell cycle
Development
Signal transduction
Other
F5 F7 F8
0
5
-log10 (p Value)
J
Factor 7
Variance explained (R²)
Methylation SE
Methylation GB
Methylation TSS
ChIP-Seq SE
mRNA
Metabolomics
Histone MS
K
Factor 7 score
L
ChIP-Seq SE
PPP1R3D
LOC389906
TBL1X
CDK5
JADE2 (1)
FLNA (1)
USP9X
AGAP3
LINC-PINT (2)
KLHL15
Scaled weight
M
Factor 8
N
Factor 8 score
O
Methylation TSS
HGF
KIT
GBP4
DKFZP434H168
GSTO1
SGK1
TNFRSF11B
NDST4
RPS6KA2
TRPC6
Scaled weight
P
+ Weights
F7 - ChIP-Seq SE
RNA binding
Cellular macromolecule catabolic process
Chromosome organization
Macromolecule catabolic process
Nuclear body
mRNA metabolic process
Negative-regulation of nucleobase/metabolic process
Cellular protein catabolic process
Cellular macromolecule localization
Chromatin organization
15.8
8.0 -log10 (p Value)
Q
+ Weights
F8 - Methylation TSS
Maintenance of protein location
Signaling receptor binding
Mucopolysaccharide metabolic process
Cellular response to chemical stress
Regulation of DNA biosynthetic process
Aortic valve stenosis
Glycosphingolipid metabolic process
DNA biosynthetic process
Cellular response to hormone stimulus
Response to hormone
3.9
8.2
-log10 (p Value)
R
r = -0.26, p = 1
r = 0.79, p = 2.34E-06
r = -0.32, p = 1
r = 0.12, p = 1
r = -0.40, p = 9.70E-01
r = -0.21, p = 1
r = 0.42, p = 7.15E-01
Methylation SE Total
ChIP Seq SE Total
Methylation GB Total
mRNA Total
Methylation TSS Total
Metabolomics Total
Histone MS Total
Factor 7
Fig S2

Figure S2. Multi-omics factor analysis (MOFA) of TNBC. Related to Figure 2.

(A) Samples and features across each dataset used for MOFA analysis.

(B) Heatmap of variance explained in each dataset (rows) by each MOFA factor (columns). Bar-graph shows total variance explained in each dataset considering all MOFA factors in the full model.

(C) Scatter plots of total signal in each dataset against Factor 2 scores. Holm-adjusted Pearson correlation test p values are shown.

(D) Scatter plots of total signal in each dataset against Factor 3 scores. Holm-adjusted Pearson correlation test p values are shown.

(E) Bar graph showing variance explained by Factor 5 in each dataset.

(F) Bee swarm plot showing Factor 5 scores in TNBC cell lines. Points are jittered along the horizontal axis for the purpose of visualization.

(G) Scaled Factor 5 weights for the promoter methylation features with the largest positive weights for this factor. Scaled weights for each factor in each dataset are derived from the weights for that factor in that dataset by linearly rescaling the values to lie between -1 and 1 (See Methods).

(H) Heatmap for ontology gene set feature set enrichment analysis based on features in the promoter DNA methylation dataset with positive Factor 5 (F5) weights.

(I) Heatmap depicting Metacore network enrichment for promoter DNA methylation features with positive Factor 5 weights (F5), H3K27ac SE features with positive Factor 7 weights (F7) and promoter methylation features with positive Factor 8 weights (F8).

(J) Bar graph showing variance explained by Factor 7 in each dataset.

(K) Bee swarm plot showing Factor 7 scores in TNBC cell lines. Points are jittered along the horizontal axis for the purpose of visualization.

(L) Scaled Factor 7 weights for the H3K27ac SE features with the largest positive weights for this factor. Scaled weights for each factor in each dataset are derived from the weights for that factor in that dataset by linearly rescaling the values to lie between -1 and 1 (See Methods).

(M) Bar graph showing variance explained by Factor 8 in each dataset.

(N) Bee swarm plot showing Factor 8 scores in TNBC cell lines. Points are jittered along the horizontal axis for the purpose of visualization.

(O) Scaled Factor 8 weights for the promoter methylation features with the largest positive weights for this factor. Scaled weights for each factor in each dataset are derived from the weights for that factor in that dataset by linearly rescaling the values to lie between -1 and 1 (See Methods).

(P) Heatmap for ontology gene set feature enrichment analysis based on features in the H3K27ac SE dataset with positive Factor 7 (F7) weights.

(Q) Heatmap for ontology gene set feature enrichment analysis based on features in the promoter methylation dataset with positive Factor 8 (F8) weights.

(R) Scatter plots of total signal in each dataset against Factor 7 scores. Holm-adjusted Pearson correlation test p values are shown.

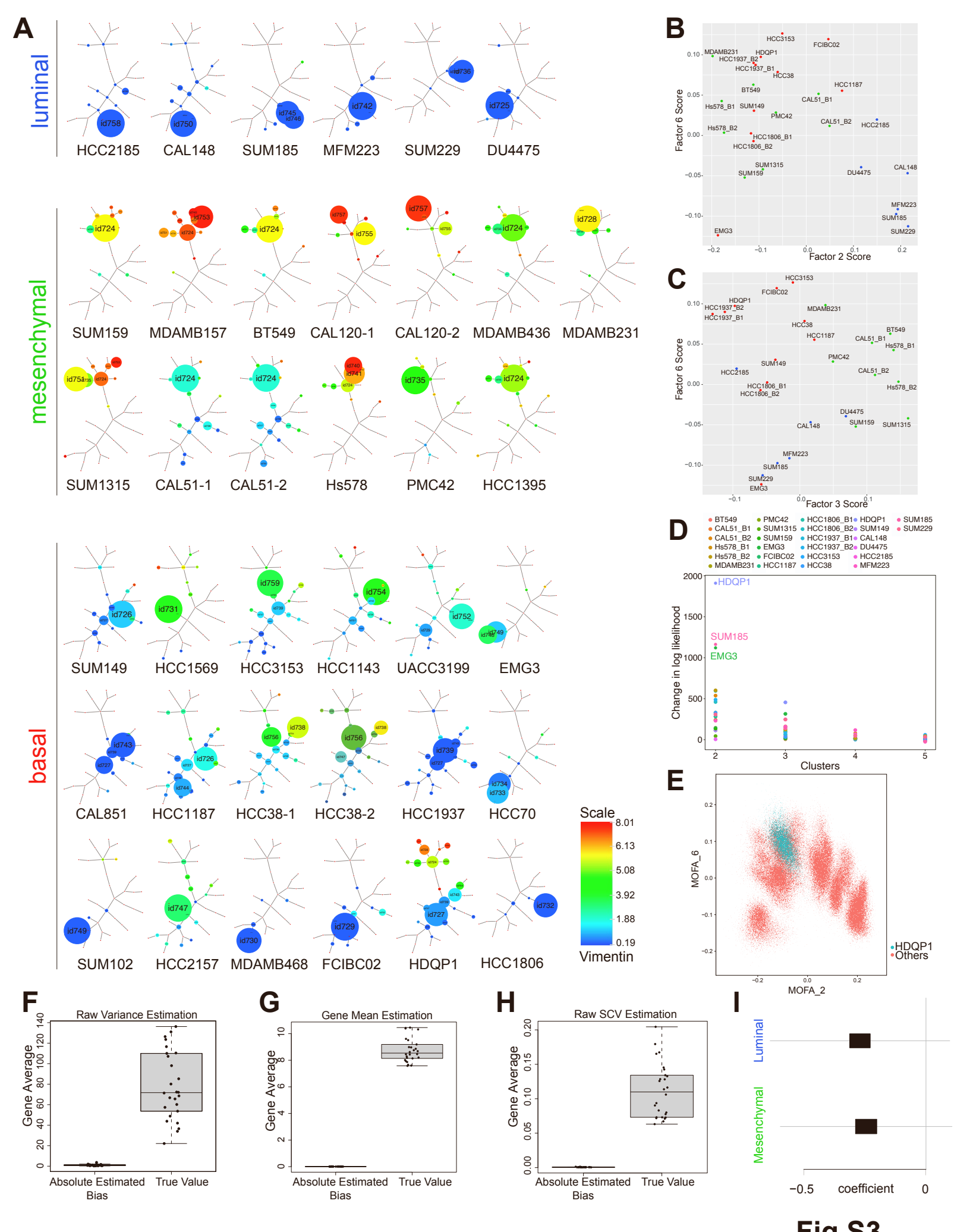
A
luminal
HCC2185
CAL148
SUM185
MFM223
SUM229
DU4475
mesenchymal
SUM159
MDAMB157
BT549
CAL120-1
CAL120-2
MDAMB436
MDAMB231
SUM1315
CAL51-1
CAL51-2
Hs578
PMC42
HCC1395
basal
SUM149
HCC1569
HCC3153
HCC1143
UACC3199
EMG3
CAL851
HCC1187
HCC38-1
HCC38-2
HCC1937
HCC70
SUM102
HCC2157
MDAMB468
FCIBC02
HDQP1
HCC1806
Scale
8.01
6.13
5.08
3.92
1.88
0.19
Vimentin
B
Factor 6 Score
Factor 2 Score
C
Factor 6 Score
Factor 3 Score
D
Change in log likelihood
Clusters
HDQP1
SUM185
EMG3
E
MOFA_6
MOFA_2
HDQP1
Others
F
Raw Variance Estimation
Gene Average
Absolute Estimated Bias
True Value
G
Gene Mean Estimation
Gene Average
Absolute Estimated Bias
True Value
H
Raw SCV Estimation
Gene Average
Absolute Estimated Bias
True Value
I
Luminal
Mesenchymal
coefficient

Fig S3

Figure S3. Single cell analyses of TNBC lines. Related to Figure 4.
(A) CyTOF-derived x-shift clustering plots for cell line samples from each TNBC type. Plots are colored by Vimentin expression in single cells from each cluster. Identifiers within plots indicate unique clusters. Three cell lines (CAL51, CAL120, and HCC38) were run in duplicates across the two batches; sample names ending in “-1” and “-2” indicate CY1 and CY2 replicates of these three cell lines, respectively.
(B) Average MOFA Factor 2 and Factor 6 scores of single cells from each sample (see Methods)
(C) Average MOFA Factor 3 and Factor 6 scores of single cells from each sample (see Methods)
(D) For each sample, the change in log likelihood compared to a model with one fewer cluster is shown for models with 2, 3, 4, or 5 clusters.
(E) Inferred MOFA Factor 2 and MOFA Factor 6 scores for HDQP1 single cells (blue) and all other single cells (red).
(F) Average absolute estimated bias for raw variance estimation across genes for parameter settings corresponding to each single-cell sample (left) and average ground truth raw variance values for each single-cell sample (right). Bias was estimated as the average difference between the ground truth value and the estimated value across 1,000 simulated datasets.
(G) Average absolute estimated bias for gene mean expression estimation across genes for parameter settings corresponding to each single-cell sample (left) and average ground truth gene mean expression value across genes for each single-cell sample (right). Bias was estimated as the average difference between the ground truth value and the estimated value across 1,000 simulated datasets.
(H) Average absolute estimated bias for raw SCV estimation across genes for parameter settings corresponding to each single-cell sample (left) and average ground truth raw SCV value across genes for each single-cell sample (right). Bias was estimated as the average difference between the ground truth value and the estimated value across 1,000 simulated datasets.
(I) Confidence interval for estimated luminal and mesenchymal coefficients in a linear mixed effects model of the logarithm of raw transcriptional variance (see Methods). The basal subtype is the base level.

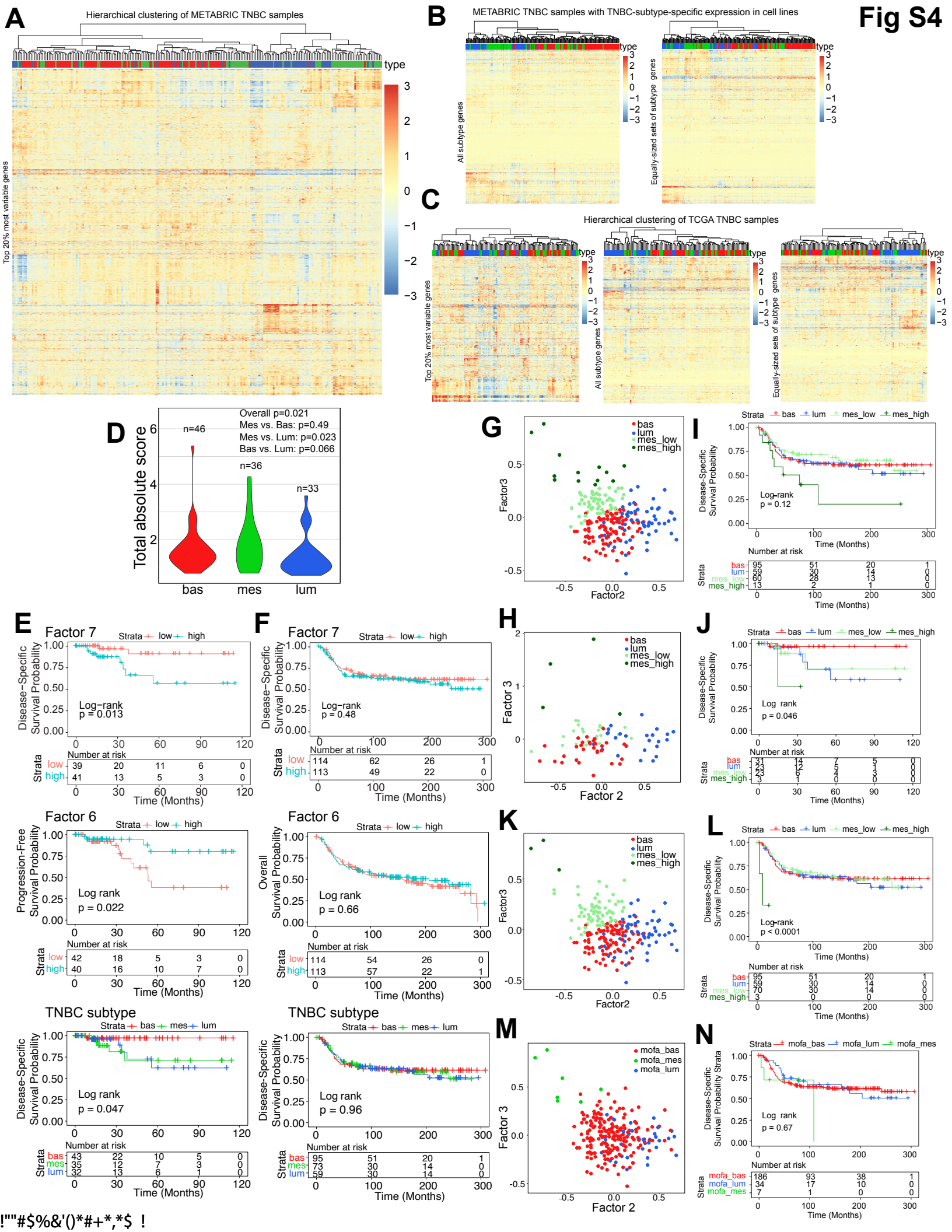
Fig S4
A
Hierarchical clustering of METABRIC TNBC samples
type
Top 20% most variable genes
B
METABRIC TNBC samples with TNBC-subtype-specific expression in cell lines
All subtype genes
Equally-sized sets of subtype genes
C
Hierarchical clustering of TCGA TNBC samples
D
Overall p=0.021
Mes vs. Bas: p=0.49
Mes vs. Lum: p=0.023
Bas vs. Lum: p=0.066
n=46
n=36
n=33
Total absolute score
bas
mes
lum
E
Factor 7
Log−rank p = 0.013
F
Factor 7
Log−rank p = 0.48
G
Factor3
Factor2
I
Log-rank p = 0.12
H
Factor 3
Factor 2
J
Log rank p = 0.046
Factor 6
Progression-Free Survival Probability
Log rank p = 0.022
Factor 6
Overall Survival Probability
Log rank p = 0.66
K
L
Log-rank p < 0.0001
TNBC subtype
Log rank p = 0.047
TNBC subtype
Log rank p = 0.96
M
mofa_bas
mofa_mes
mofa_lum
Factor 3
Factor 2
N
Log rank p = 0.67
Disease-Specific Survival Probability
Time (Months)
Number at risk
Strata

!""#$%&'()*#+*,*$!

Figure S4. Clinical relevance of TNBC subtype heterogeneity. Related to Figure 1.

(A) Results of unbiased hierarchical clustering of METABRIC TNBC samples based on expression data, using the top 20% most variable genes. Annotation track represents assigned subtype using TNBC subtype signatures defined from cell lines. Values are capped at +/-3 for the purpose of visualization.

(B) Hierarchical clustering of METABRIC TNBC samples using all genes with TNBC-subtype-specific expression in cell lines (See Table S1) (left), and equally sized sets of genes with subtype-specific expression in cell lines (right). Values are capped at +/-3 for the purpose of visualization.

(C) Heatmaps showing clustering of TCGA TNBC sample using top 20% most variable genes (left), all genes with TNBC-subtype-specific expression in cell lines (See Table S1) (middle), and equally sized sets with subtype-specific expression in cell lines (right).

(D) Total absolute CIBERSORTx scores for TCGA samples assigned to each of the three TNBC types. Overall p=0.021, Kruskal-Wallis test; pairwise p values (p=0.49 mes vs bas, p=0.023 mes vs lum, p=0.066 bas vs lum from Dunn's test, adjusted using Holm's method).

(E) Kaplan-Meier survival curves of patients in the TCGA cohort split by median Factor 7 score (top), median Factor 6 score (middle), and assigned TNBC subtype (bottom).

(F) Kaplan-Meier survival curves of patients in the METABRIC cohort split by median Factor 7 score (top), median Factor 6 score (middle), and assigned TNBC subtype (bottom).

(G) Scatter plot of inferred Factor 2 and 3 scores for METABRIC TNBC samples, colored by refined TNBC subtype.

(H) Scatter plot of inferred Factor 2 and 3 scores for TCGA TNBC samples, colored by refined TNBC subtype.

(I) METABRIC-based Kaplan-Meier survival curves for samples split by refined TNBC subtype.

(J) TCGA-based Kaplan-Meier survival curves for samples split by refined TNBC subtype.

(K) Scatter plot of inferred Factor 2 and 3 values for METABRIC TNBC samples, colored by refined TNBC subtype, using an alternative clustering-based approach to define the refined subtypes.

(L) METABRIC-based Kaplan-Meier survival curves for samples split by refined TNBC subtype, using an alternative clustering-based approach to define the refined subtypes.

(M) Scatter plot of inferred Factor 2 and 3 values for METABRIC TNBC samples, colored by TNBC subtype using an alternative clustering-based approach to define the TNBC subtypes.

(N) METABRIC-based Kaplan-Meier survival curves for samples split by TNBC subtype, using an alternative clustering-based approach to define the TNBC subtypes.

Fig S5

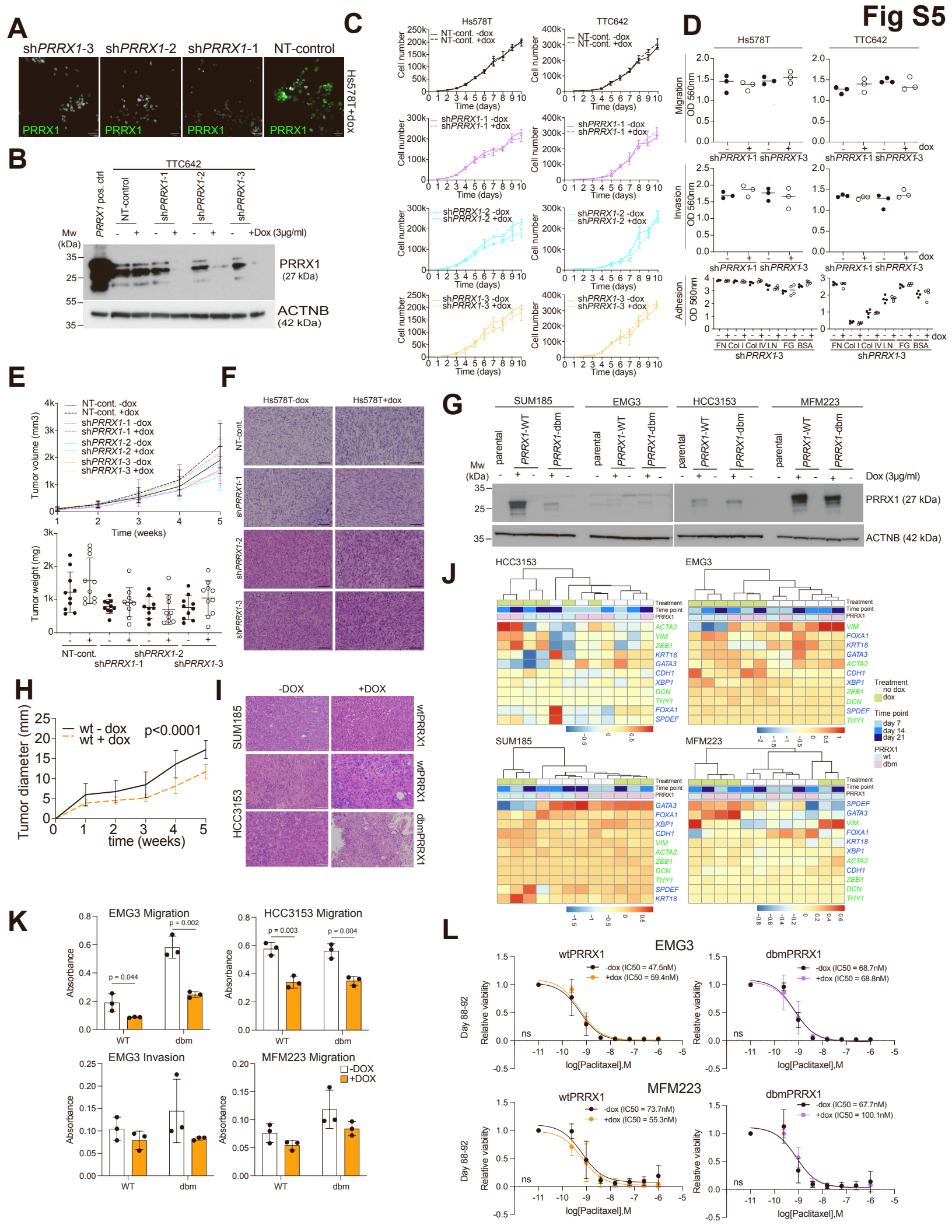
A
shPRRX1-3
shPRRX1-2
shPRRX1-1
NT-control
Hs578T+dox
PRRX1
B
TTC642
PRRX1 pos. ctrl
NT-control
shPRRX1-1
shPRRX1-2
shPRRX1-3
Mw (kDa)
+Dox (3μg/ml)
PRRX1 (27 kDa)
ACTNB (42 kDa)
C
Hs578T
TTC642
NT-cont. -dox
NT-cont. +dox
shPRRX1-1 -dox
shPRRX1-1 +dox
shPRRX1-2 -dox
shPRRX1-2 +dox
shPRRX1-3 -dox
shPRRX1-3 +dox
Cell number
Time (days)
D
Migration OD 560nm
Invasion OD 560nm
Adhesion OD 560nm
dox
FN Col I Col IV LN FG BSA
E
Tumor volume (mm3)
Time (weeks)
Tumor weight (mg)
NT-cont.
shPRRX1-2
shPRRX1-1
shPRRX1-3
F
Hs578T-dox
Hs578T+dox
G
SUM185
EMG3
HCC3153
MFM223
parental
PRRX1-WT
PRRX1-dbm
Dox (3μg/ml)
PRRX1 (27 kDa)
ACTNB (42 kDa)
H
wt - dox
wt + dox
p<0.0001
Tumor diameter (mm)
time (weeks)
I
-DOX
+DOX
SUM185
HCC3153
wtPRRX1
dbmPRRX1
J
HCC3153
EMG3
SUM185
MFM223
Treatment
Time point
PRRX1
ACTA2
VIM
ZEB1
KRT18
GATA3
CDH1
XBP1
DCN
THY1
FOXA1
SPDEF
no dox
dox
day 7
day 14
day 21
wt
dbm
K
EMG3 Migration
HCC3153 Migration
EMG3 Invasion
MFM223 Migration
Absorbance
p = 0.044
p = 0.002
p = 0.003
p = 0.004
WT
dbm
-DOX
+DOX
L
EMG3
MFM223
wtPRRX1
dbmPRRX1
-dox (IC50 = 47.5nM)
+dox (IC50 = 59.4nM)
-dox (IC50 = 68.7nM)
+dox (IC50 = 68.8nM)
-dox (IC50 = 73.7nM)
+dox (IC50 = 55.3nM)
-dox (IC50 = 67.7nM)
+dox (IC50 = 100.1nM)
Day 88-92
Relative viability
log[Paclitaxel],M
ns

Figure S5. Characterization of PRRX1. Related to Figure 5.

(A) Immunofluorescence analyses of PRRX1 protein expression in Hs578T cells, respectively expressing three independent TET-doxycycline inducible *PRRX1*-targeting shRNAs as well as non-targeting control. Scale bar corresponds to 50μm.

(B) Immunoblot analyses of PRRX1 protein expression in TTC642 cells expressing three independent TET-doxycycline inducible *PRRX1*-targeting shRNAs as well as non-targeting control.

(C) Proliferation of Hs578T and TTC642 cell lines after downregulation of *PRRX1* by TET-inducible shRNAs. Error bars represent SEM; n=3 replicates.

(D) Migration, invasion, and cell adhesion assays in Hs578T and TTC642 cell lines after downregulation of *PRRX1* by TET-inducible shRNAs**.** Error bars represent SD, n=3 replicates.

(E) Tumor weights and volumes of xenografts derived from Hs578T cell line expressing TET-inducible shRNAs with and without doxycycline in diet. Error bars represent SEM, n=10/group.

(F) Hematoxylin and eosin-stained images of xenografts of Hs578T TNBC line expressing TET-inducible sh*PRRX1* from mice with and without doxycycline in diet. Scale bar corresponds to 50μm.

(G) Immunoblot analysis for PRRX1 protein expression in basal (HCC3153, EMG3) and luminal (SUM185, MFM223) lines expressing doxycycline inducible wt and dbm PRRX1 protein.

(H) Plot depicting change in tumor diameter over time of xenografts derived from SUM185 cell line expressing wt PRRX1. Each data bar represents the mean tumor diameter of 10 tumors; error bar represents SEM ($P < 0.0001$ for a two-tailed Student's t-test).

(I) Hematoxylin and eosin-stained images of xenografts of HCC3153 and SUM185 cell line exogenously expressing WT or MUT *PRRX1* from mice with and without doxycycline in diet. Scale bar 100μm.

(J) Heat maps depicting clustering of the samples based on the expression of known luminal (blue) and mesenchymal (green) genes.

(K) Migration and invasion assays in the indicated cell lines after induction of wt or dbm *PRRX1* by doxycycline (DOX)**.** Error bars represent SEM (EMG3: $p=0.044$ (wt) and $p=0.002$ (dbm); HCC3153: $p=0.003$ (wt) and $p=0.004$ (dmb) values for a two-tailed Student's t-test), n=3 replicates.

(L) Plots depicting viable cell numbers of EMG3 and MFM223 cells following paclitaxel treatment and induction of wt or dbm PRRX1 expression by doxycycline for the indicated days. Error bars represent SEM (all p values not significant for a nonlinear fit test), n=3 replicates.

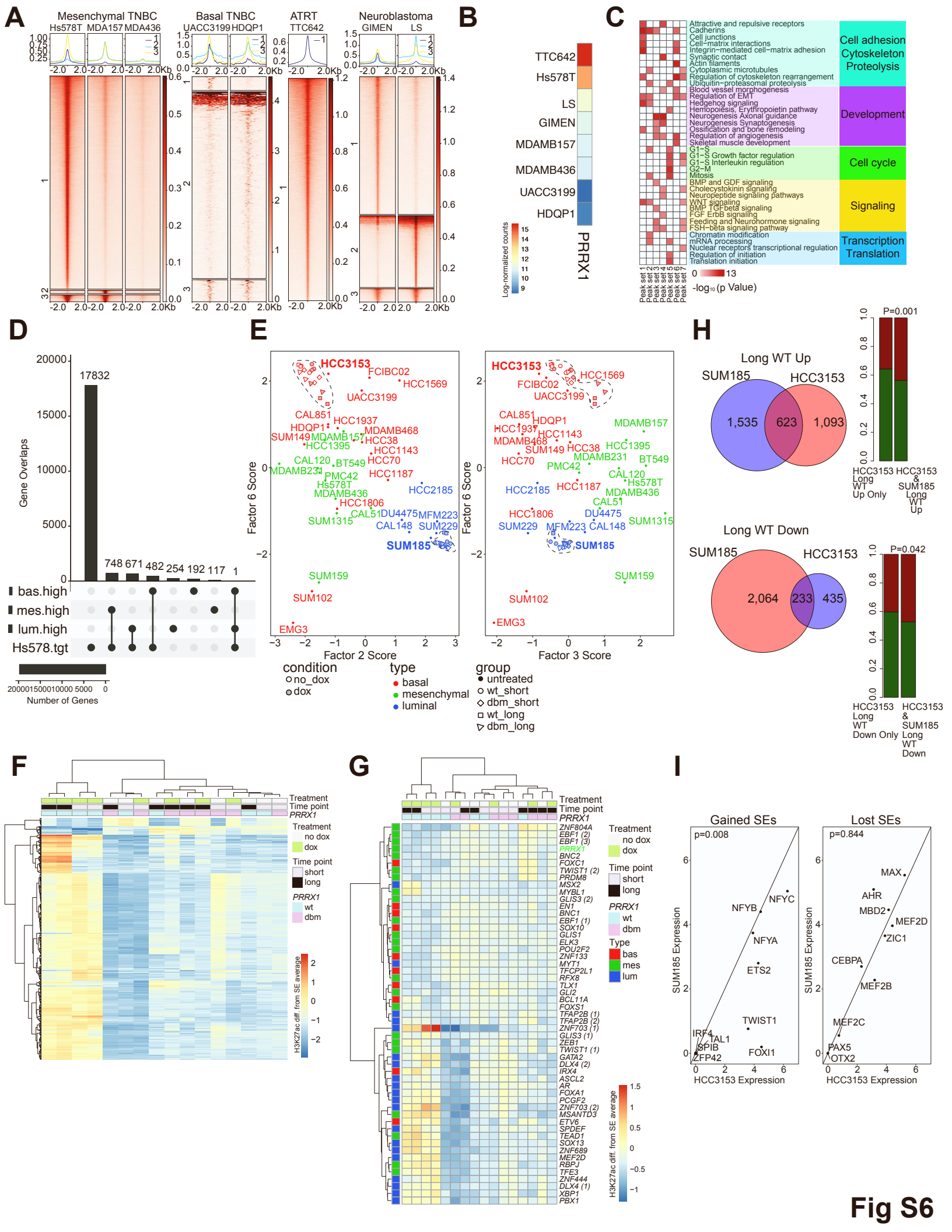
A
Mesenchymal TNBC
Hs578T MDA157 MDA436
Basal TNBC
UACC3199 HDQP1
ATRT
TTC642
Neuroblastoma
GIMEN LS
B
TTC642
Hs578T
LS
GIMEN
MDAMB157
MDAMB436
UACC3199
HDQP1
PRRX1
Log-normalized counts
C
Attractive and repulsive receptors
Cadherins
Cell junctions
Cell−matrix interactions
Integrin−mediated cell−matrix adhesion
Synaptic contact
Actin filaments
Cytoplasmic microtubules
Regulation of cytoskeleton rearrangement
Ubiquitin−proteasomal proteolysis
Blood vessel morphogenesis
Regulation of EMT
Hedgehog signaling
Hemopoiesis; Erythropoietin pathway
Neurogenesis Axonal guidance
Neurogenesis Synaptogenesis
Ossification and bone remodeling
Regulation of angiogenesis
Skeletal muscle development
G1−S
G1−S Growth factor regulation
G1−S Interleukin regulation
G2−M
Mitosis
BMP and GDF signaling
Cholecystokinin signaling
Neuropeptide signaling pathways
WNT signaling
BMP TGFbeta signaling
FGF ErbB signaling
Feeding and Neurohormone signaling
FSH−beta signaling pathway
Chromatin modification
mRNA processing
Nuclear receptors transcriptional regulation
Regulation of initiation
Translation initiation
Cell adhesion Cytoskeleton Proteolysis
Development
Cell cycle
Signaling
Transcription Translation
Peak set 1
Peak set 2
Peak set 3
Peak set 4
Peak set 5
Peak set 6
Peak set 7
-log10(p Value)
D
Gene Overlaps
17832
748 671 482 254 192 117 1
bas.high
mes.high
lum.high
Hs578.tgt
Number of Genes
E
Factor 6 Score
Factor 2 Score
Factor 3 Score
HCC3153
FCIBC02
HCC1569
UACC3199
CAL851
HCC1937
HDQP1
MDAMB468
SUM149
MDAMB157
HCC38
HCC1395
HCC1143
CAL120
BT549
HCC70
MDAMB231
PMC42
HCC1187
Hs578T
MDAMB436
HCC2185
HCC1806
CAL51
DU4475
MFM223
SUM1315
CAL148
SUM229
SUM185
SUM159
SUM102
EMG3
condition
no_dox
dox
type
basal
mesenchymal
luminal
group
untreated
wt_short
dbm_short
wt_long
dbm_long
H
Long WT Up
SUM185
HCC3153
1,535
623
1,093
P=0.001
HCC3153 Long WT Up Only
HCC3153 & SUM185 Long WT Up
Long WT Down
2,064
233
435
P=0.042
HCC3153 Long WT Down Only
HCC3153 & SUM185 Long WT Down
F
Treatment
Time point
PRRX1
no dox
dox
short
long
wt
dbm
H3K27ac diff. from SE average
G
ZNF804A
EBF1 (2)
EBF1 (3)
PRRX1
BNC2
FOXC1
TWIST1 (2)
PRDM8
MSX2
MYBL1
GLIS3 (2)
EN1
BNC1
EBF1 (1)
SOX10
GLIS1
ELK3
POU2F2
ZNF133
MYT1
TFCP2L1
RFX8
TLX1
GLI2
BCL11A
FOXS1
TFAP2B (1)
TFAP2B (2)
ZNF703 (1)
GLIS3 (1)
ZEB1
TWIST1 (1)
GATA2
DLX4 (2)
IRX4
ASCL2
AR
FOXA1
PCGF2
ZNF703 (2)
MSANTD3
ETV6
SPDEF
TEAD1
SOX13
ZNF689
MEF2D
RBPJ
TFE3
ZNF444
DLX4 (1)
XBP1
PBX1
Type
bas
mes
lum
I
Gained SEs
p=0.008
Lost SEs
p=0.844
SUM185 Expression
HCC3153 Expression
NFYB
NFYC
NFYA
ETS2
TWIST1
IRF4
TAL1
SPIB
ZFP42
FOXI1
MAX
AHR
MBD2
MEF2D
ZIC1
CEBPA
MEF2B
MEF2C
PAX5
OTX2

Fig S6

Figure S6. PRRX1-mediated expression and chromatin changes.
(A) PRRX1 ChIP-seq peaks across mesenchymal and basal TNBC, rhabdoid (ATRT), and neuroblastoma cell lines. All combinations represented in 1% of peaks in each type of cell line are shown (Mesenchymal plot: 1: Hs578T (80,757 peaks), 2: MDAMB157 (1,188 peaks), 3: Hs578T and MDAMB157 (2,617 peaks), Basal plot: 1: UACC3199 (29 peaks), 2: HDQP1 (391 peaks), 3: UACC3199 and HDQP1 (46 peaks), Rhabdoid plot: 1: TTC642 (46,296 peaks), neuroblastoma plot: 1: GIMEN (16,187 peaks), 2: LS (8,437 peaks), 3: GIMEN and LS (1,566 peaks)). The color scheme represents PRRX1 ChIP-seq signal score +/- 2,000 bp from the peak center in each cell line. For each cell line data is shown from the ChIP-seq experiment which yielded a larger number of peaks.
(B) PRRX1 mRNA levels in the cell lines used for ChIP-seq.
(C) Metacore network enrichment analysis for shared and unique PRRX1 peaks across TNBC, neuroblastoma and rhabdoid tumor cell line shown in Fig. 5C.
(D) Number of genes specifically expressed in each TNBC sub-type overlapping the PRRX1 Hs578T target set (see Methods). Top right inset bar plot shows the total size of each gene set.
(E) MOFA Factor 2 and Factor 6 scores (left) and MOFA Factor 3 and Factor 6 scores (right) for each sample, calculated based on super-enhancer H3K27ac signal (see Methods). Red and blue outlined shapes within dotted lines represent HCC3153 and SUM185 samples from the PRRX1 over-expression H3K27ac experiment respectively.
(F) Heatmap showing clustering of SUM185 PRRX1 over-expression samples and corresponding controls based on H3K27ac signal in the top 20% most variable super-enhancers.
(G) Heatmap of super-enhancer H3K27ac signal for TNBC subtype-specific transcription factors in SUM185 PRRX1 over-expression samples and corresponding controls.
(H) Venn diagrams: Overlaps between positively and negative differentially acetylated super-enhancer regions under long-term wt PRRX1 over-expression in HCC3153 and SUM185. Bar plots: Proportion of HCC3153-unique and shared (between HCC3153 and SUM185) long-term changed superenhancers in HCC3153 that were classified as indirect (red) or direct changes (green) in HCC3153.
(I) Expression of putative PRRX1 co-binding TFs in SUM185 and HCC3153. Left and right plots show transcription factors identified from long-term HCC3153 gained and lost regions, respectively. Black diagonal lines are shown for reference and indicate equal expression in the two cell lines. P=0.008 (gained SEs) and p=0.844 (lost SEs) for Mann-Whitney *U* test.

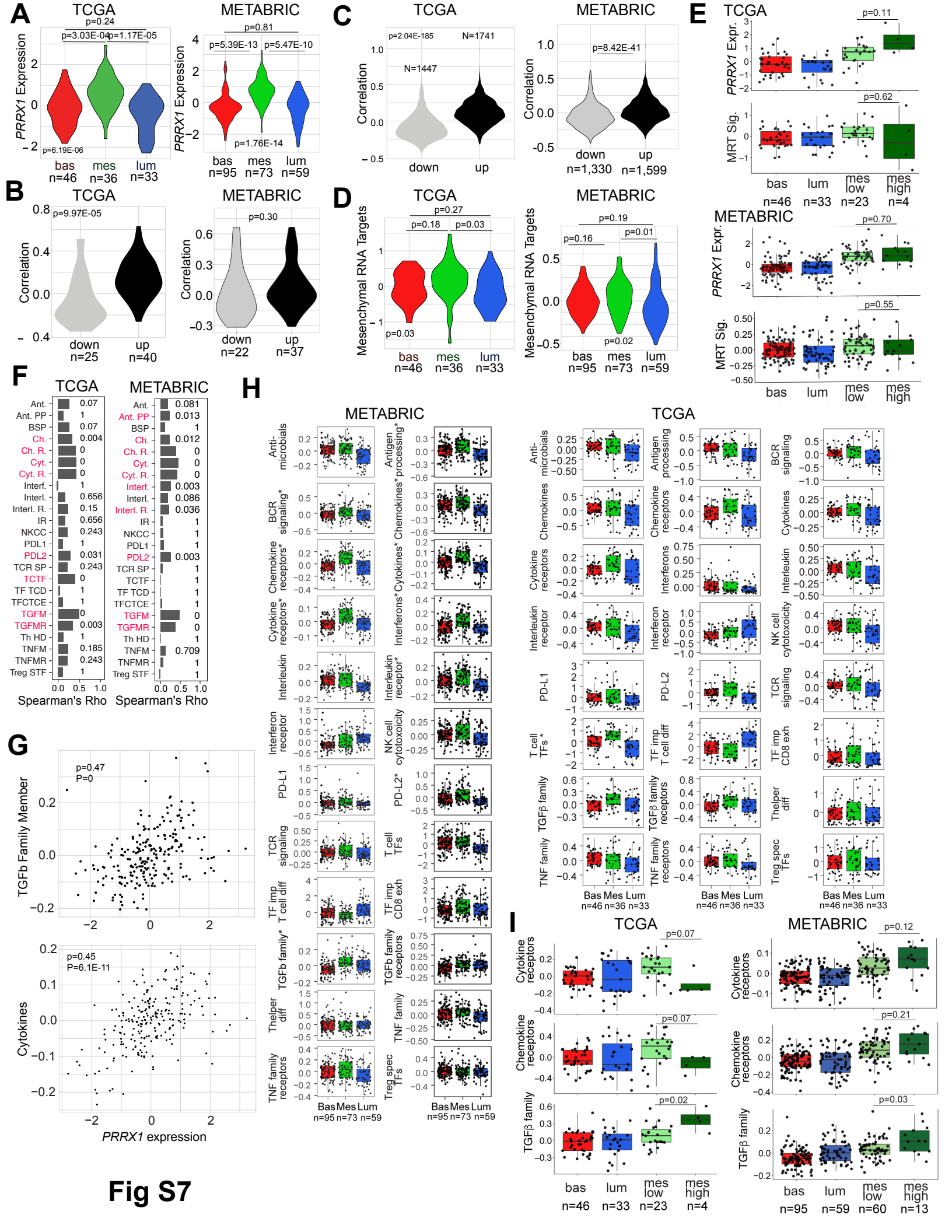
A
TCGA
METABRIC
PRRX1 Expression
p=0.24
p=3.03E-04
p=1.17E-05
p=6.19E-06
p=0.81
p=5.39E-13
p=5.47E-10
p=1.76E-14
bas n=46
mes n=36
lum n=33
bas n=95
mes n=73
lum n=59
B
TCGA
METABRIC
Correlation
p=9.97E-05
p=0.30
down n=25
up n=40
down n=22
up n=37
C
TCGA
METABRIC
Correlation
p=2.04E-185
N=1741
N=1447
p=8.42E-41
down
up
down n=1,330
up n=1,599
D
TCGA
METABRIC
Mesenchymal RNA Targets
p=0.27
p=0.18
p=0.03
p=0.03
p=0.19
p=0.16
p=0.01
p=0.02
bas n=46
mes n=36
lum n=33
bas n=95
mes n=73
lum n=59
E
TCGA
PRRX1 Expr.
MRT Sig.
p=0.11
p=0.62
bas n=46
lum n=33
mes low n=23
mes high n=4
METABRIC
PRRX1 Expr.
MRT Sig.
p=0.70
p=0.55
bas
lum
mes low
mes high
F
TCGA
METABRIC
Ant.
Ant. PP
BSP
Ch.
Ch. R.
Cyt.
Cyt. R.
Interf.
Interl.
Interl. R.
IR
NKCC
PDL1
PDL2
TCR SP
TCTF
TF TCD
TFCTCE
TGFM
TGFMR
Th HD
TNFM
TNFMR
Treg STF
Spearman's Rho
Spearman's Rho
G
p=0.47
P=0
TGFb Family Member
p=0.45
P=6.1E-11
Cytokines
PRRX1 expression
H
METABRIC
TCGA
Anti-microbials
Antigen processing
BCR signaling*
Chemokines*
Chemokine receptors*
Cytokines*
Cytokine receptors*
Interferons*
Interleukin
Interleukin receptor*
Interferon receptor
NK cell cytotoxicity
PD-L1
PD-L2*
TCR signaling
T cell TFs
TF imp T cell diff
TF imp CD8 exh
TGFb family*
TGFb family receptors
Thelper diff
TNF family
TNF family receptors
Treg spec TFs
Bas Mes Lum
n=95 n=73 n=59
Anti-microbials
Antigen processing
BCR signaling
Chemokines
Chemokine receptors
Cytokines
Cytokine receptors
Interferons
Interleukin
Interleukin receptor
Interferon receptor
NK cell cytotoxicity
PD-L1
PD-L2
TCR signaling
T cell TFs *
TF imp T cell diff
TF imp CD8 exh
TGFβ family
TGFβ family receptors
Thelper diff
TNF family
TNF family receptors
Treg spec TFs
Bas Mes Lum
n=46 n=36 n=33
I
TCGA
METABRIC
Cytokine receptors
Chemokine receptors
TGFβ family
p=0.07
p=0.07
p=0.02
p=0.12
p=0.21
p=0.03
bas n=46
lum n=33
mes low n=23
mes high n=4
bas n=95
lum n=59
mes low n=60
mes high n=13

Fig S7

Figure S7. Clinical relevance of PRRX1. Related to Figure 5.

(A) Mean-centered *PRRX1* expression in TCGA and METABRIC samples assigned to each of the three TNBC types. Holm-adjusted Dunn test p values are shown. PRRX1 expression differs among subtypes (p=6.2E-6 (TCGA); p=1.76E-14 (METABRIC) for Kruskal-Wallis test).

(B) Pearson correlation with *PRRX1* expression for putatively positively and negatively regulated genes in the 'Mesenchymal RNA target' (MRT) gene set for TCGA and METABRIC TNBC samples. Up, positive targets; down, negative targets. Mann-Whitney *U* test p values are shown.

(C) Pearson correlation with *PRRX1* expression for putatively positively and negatively regulated genes in the 'Hs578T RNA target' (HsRT) gene set for TCGA and METABRIC TNBC samples. Up, positive targets; down, negative targets. Mann-Whitney *U* test p values are shown.

(D) Mesenchymal RNA target signature scores in TCGA and METABRIC samples assigned to each TNBC type. Overall p value from Kruskall-Wallis test; pairwise p values from Dunn's test, adjusted using Holm's method.

(E) Mean-centered *PRRX1* expression (top panels) and Mesenchymal RNA target signature score (bottom panels) in TCGA and METABRIC samples assigned to each of the four refined TNBC types. Mann-Whitney U p values are shown.

(F) Correlations between immune signature scores and PRRX1 expression in the TCGA (n=115) and METABRIC (n=214) cohorts. Samples with tied immune signature scores or PRRX1 expression values were removed for the purposes of the correlation test. Holm-adjusted p value from Spearman correlation tests are shown. Signatures with significant adjusted p values (<0.05) are colored red. Ant.:Anitmicrobials, Ant. PP:Antigen_Processing_and_Presentation, BSP: BCRSignalingPathway, Ch.: Chemokines, Ch. R.:Chemokine_Receptors, Cyt.: Cytokines, Cyt. R.:Cytokine_Receptors, Interf.: Interferons, Interl.:Interleukins, Interl. R.:Interleukins_Receptor, IR:Interferon_Receptor, NKCC: NaturalKiller_Cell_Cytotoxicity, TCR SP:TCRsignalingPathway, TCTF:T_cell_TF, TF TCD: TF_important_for_T_celll_differentiation, TFCTCE:TF_important_for_CD8_T_cell_exhaustion, TGFM: TGFb_Family_Member, TGFMR:TGFBb_Family_Member_Receptor, Th HD:Th_helper_diff, TNFM: TNF_Family_Members, TNFMR:TNF_Family_Members_Receptors, Treg STF:Treg_specific_TF.

(G) Scatter plots of TGF-β Family Member signature score (top) and Cytokines signature score (bottom) and mean-centered *PRRX1* expression in the METABRIC cohort. ρ: Spearman's rho, P: Holm-adjusted Spearman correlation test p value. Samples with tied immune signature expression values or PRRX1 expression values were removed for the purposes of the correlation test. The trend line is omitted since the relationship between *PRRX1* expression and signature expression is not assumed to be linear. n=227 (scatter plot), n=214 (correlation test).

(H) Immune signature scores in TCGA and METABRIC samples assigned to each TNBC type. *: Significant differences in scores between subtypes (Holm-adjusted Kruskal-Wallis test p value < 0.05) with significantly higher scores in mesenchymal TNBC than luminal or basal TNBC (Dunn test p value < 0.05, both comparisons). Ant.: Anitmicrobials, Ant. PP: Antigen_Processing_and_Presentation, BSP:BCRSignalingPathway, Ch.: Chemokines, Ch. R.: Chemokine_Receptors, Cyt.: Cytokines, Cyt. R.:Cytokine_Receptors, Interf.:Interferons, Interl.:Interleukins, Interl. R.:Interleukins_Receptor, IR:Interferon_Receptor, NKCC:NaturalKiller_Cell_Cytotoxicity, TCR SP:TCRsignalingPathway, TCTF:T_cell_TF, TF TCD: TF_important_for_T_celll_differentiation, TFCTCE:TF_important_for_CD8_T_cell_exhaustion, TGFM: TGFb_Family_Member, TGFMR: TGFBb_Family_Member_Receptor, Th HD:Th_helper_diff, TNFM:TNF_Family_Members,TNFMR:TNF_Family_Members_Receptors,TregSTF:Treg_specific_T.

(I) TGF-β Family Member, Chemokine Receptors, and Cytokine Receptors signature scores in TCGA and METABRIC samples assigned to each of the four refined TNBC types. Mann-Whitney *U* test p values are shown.