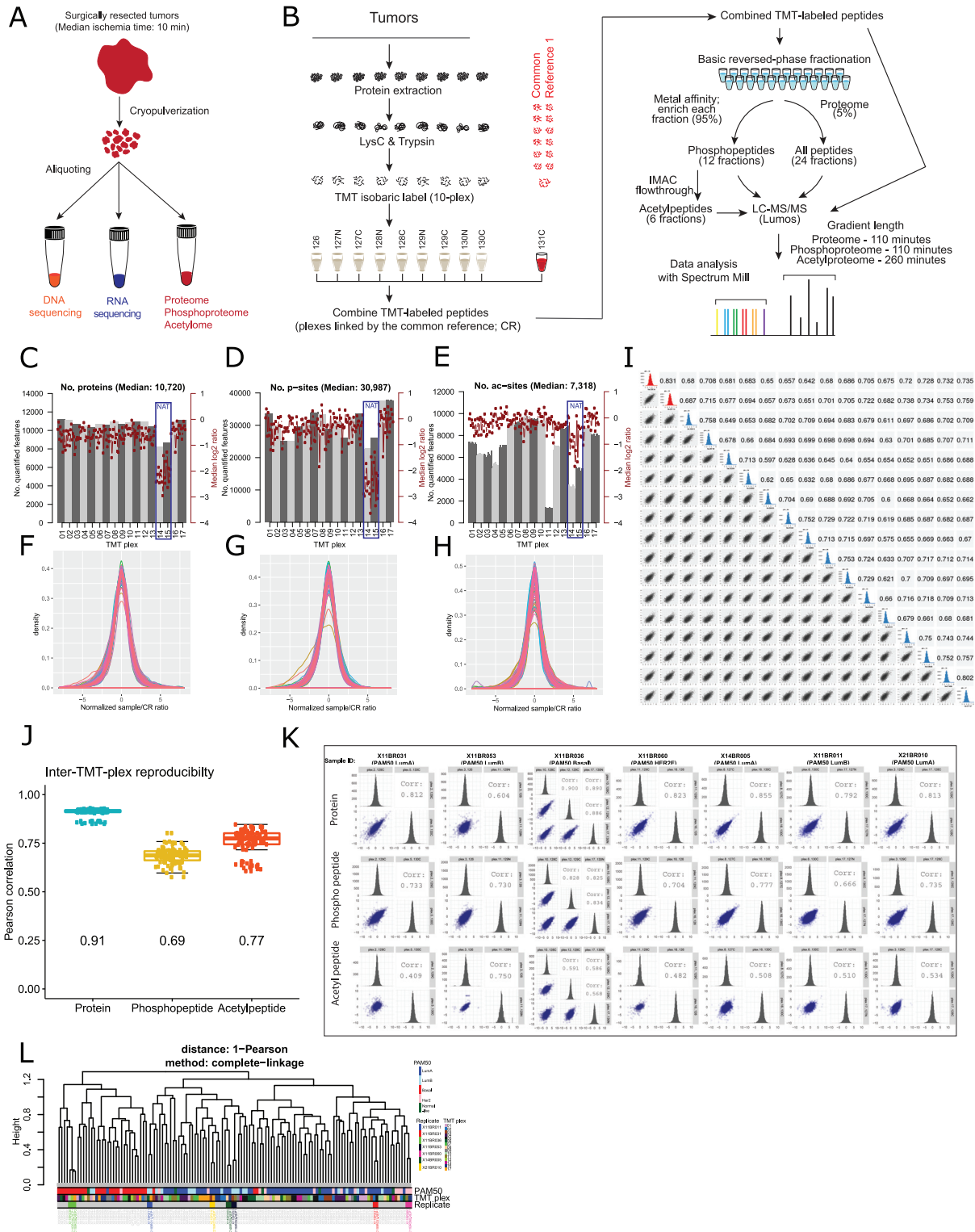# Supplemental Figures

Figure S1. Experimental Workflow and Data Quality Metrics, Related to Figure 1

A) Schematic representation of sample processing steps. Fresh frozen tumors were cryopulverized, aliquoted and subjected to genomics and proteomics analyses. Cryopulverization promoted sample uniformity across multi-omics analyses.

B) Schematic representation of the workflow used for proteome, phosphoproteome and acetylproteome analyses. Tandem mass tags (TMT) were used to multiplex 9 tumors and the common reference in a single TMT10-plex. The common reference was constructed from 40 tumors based on hormone receptor status, including 9 triple negative, 12 HER2 positive, and 19 estrogen receptor positive specimens, and was used to connect the TMT-plexes (STAR Methods), enabling consistent identification and reproducible quantification of proteins, phosphorylation and acetylation sites.

C) Bar chart depicting the number of quantified proteins per channel per TMT plex. Alternating blocks represent all channels within the indicated plex. Red points and connecting lines show the median log2 ratio per channel (right axis). The blue boxes highlight plexes 14 and 15, which comprised normal adjacent tissue (NAT) and were excluded from further analysis chiefly for biological but also for QC reasons (STAR Methods).

D) Same layout as C) showing phosphosites.

E) Same layout as C) showing acetylation sites. Consistent with the low acetyl-enrichment yield derived from plex 11 (STAR Methods, "PSM quality control and Quantification using TMT ratios"), the number of quantified acetylation sites was significantly lower in this plex.

F) Profile plots of two-component normalized protein / common reference (CR) ratios.

G) Profile plots of phosphosite / CR ratios.

H) Profile plots of acetylsite / CR ratios.

I) Pairwise inter-plex correlations between phosphopeptide-level log reporter ion intensities of the common reference channel (131), measured in all 17 TMT-10 plexes. Lower triangle: pairwise scatterplots; diagonal: histograms depicting distribution of log intensities; upper triangle: Pearson correlation calculated on pairwise complete observations of log-transformed intensities. Red histograms depict TMT plexes 14 and 15, which contained normal adjacent tissue (NAT) and were excluded from the datasets used in the manuscript. Blue histograms correspond to 15 tumor plexes. Average inter-plex Pearson correlation on the phosphopeptide level across all tumor plexes was 0.69.

J) Boxplots summarizing pairwise inter-plex Pearson correlations between log reporter ion intensities of the common reference channel (131) across 15 tumor TMT-plexes on protein, phosphopeptide and acetylpeptide levels. Boxes depict the interquartile range (IQR) of the scores with horizontal lines depicting the median. Whiskers extend to 1.5xIQR from Q1 (25th percentile) and Q3 (75th percentile), respectively. Points depict individual data points.Isobaric-label LC-MS/MS experiments are purpose-built for a ratio-based quantification strategy. After samples are mixed to form a plex, the peptides from all samples within that plex should experience the same effects due to subsequent experimental sources of variation (offline basic-pH reversed-phase fractionation, phosphopeptide or acetylpeptide enrichment, on-line reversed-phase fractionation, selection of precursor ions for MS/MS, and mass spectrometer signal strength). While these collective effects may lead to substantial inter-plex variation in peptide-level reporter ion intensities, variation of the peptide-level ratios of sample to common reference can be expected to be consistent (though subject to peptide-level missing values between plexes due to individual peptides not always being selected for MS/MS).

K) Pairwise Spearman correlations between two-component normalized sample/CR ratios of replicates scattered across ten different TMT-plexes.

L) Dendrogram illustrating the clustering of two-component normalized sample/CR phosphopeptide-level ratios based on all quantified phosphosites without additional filtering (e.g., for missing values, most variable sites, etc.; n = 63,360). The annotation tracks below the leaves of the dendrogram depict the PAM50 subtype, the TMT10-plex and color-coordinated replicate measurements (gray if no replicate measurements). Samples did not cluster by TMT10 plex, but chiefly by PAM50 subtype. Replicates measured in different TMT-plexes acquired on the mass spectrometer months apart cluster together. Dendrogram was derived from complete-linkage hierarchical clustering using 1-Pearson correlation as the distance metric.
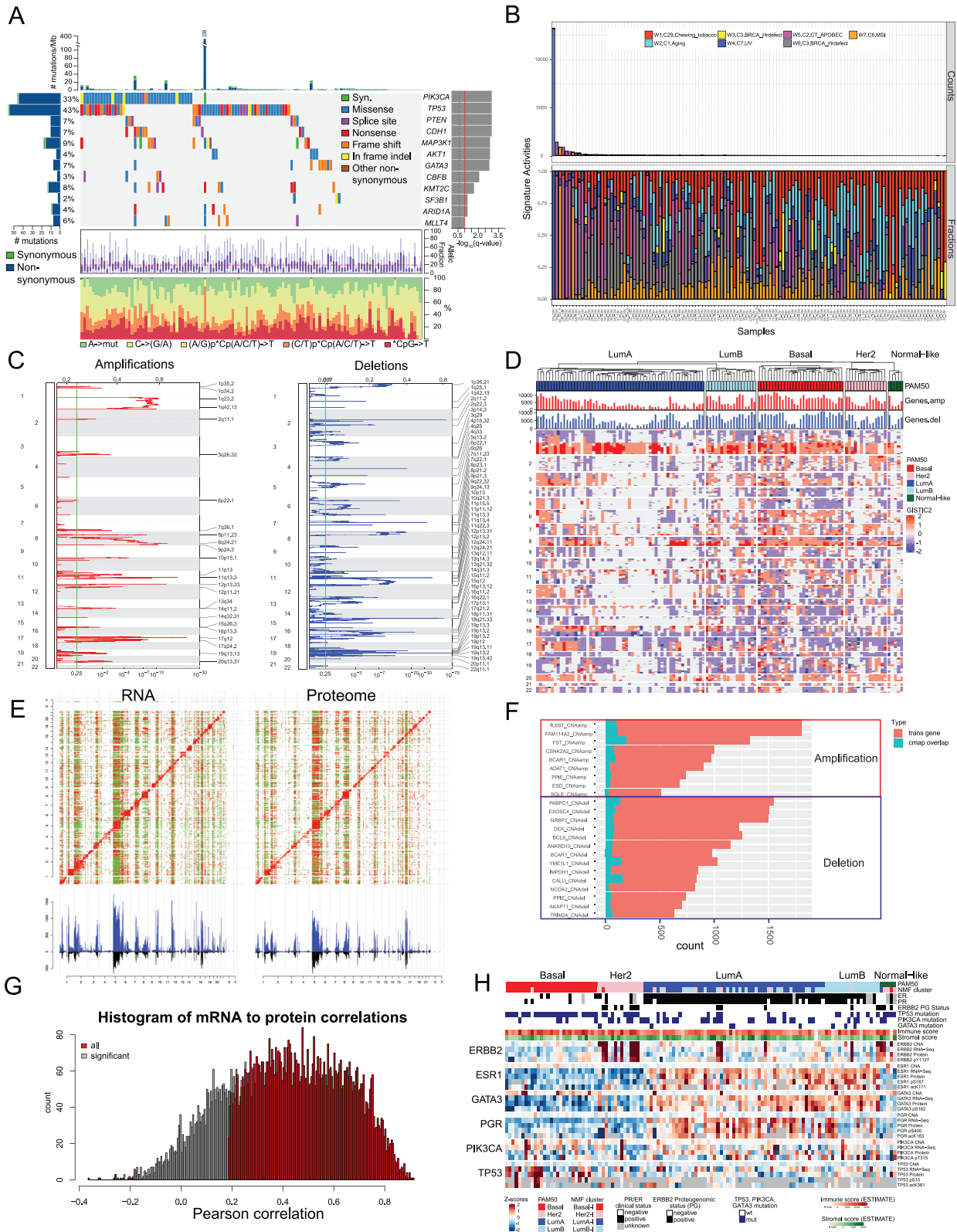
(legend on next page)

Figure S2. PG Landscape, Related to Table 1

A) CoMut (co-occurrence of mutations) plot summarizing the mutational landscape of this cohort. Each column represents a tumor sample. The top histogram shows the mutational burden of each tumor. The matrix shows mutation types of significantly (q < 0.1) mutated genes (SMG) identified by MutSig2CV. Plots below the matrix visualize the allelic fraction and substitution frequencies of somatic mutations. The bar chart to the left indicates the percentage of samples with each SMG, mutation number and type (synonymous or non-synonymous); the one to the right shows the significance of that SMG in the cohort.

B) Mutational signatures detected in this cohort. Trinucleotide sequence motif: seven distinct mutational signatures were identified (Table 1).

C) Copy number amplifications and deletions analysis. The left panel depicts genomic positions of amplified regions, with x-axes representing the normalized amplification signals (top) and significance by Q value (bottom). The right panel shows genomic positions of deleted regions, with X-axes representing the normalized deletion signals (top) and significance by Q value (bottom). The green lines represent the significance cutoff at Q value = 0.25.

D) Heatmap depicting all detected somatic copy number aberrations detected in this cohort. Genes are ordered by chromosomal location (y axis) and samples are organized by PAM50 subtype and separately clustered for each subtype (x axis).

E) Correlations of SNCA (x axis) to mRNA (left) and protein (right) abundances (y axis).

F) Results of the CMAP analysis. Potential driver genes (permutation FDR < 0.26) of observed CNA correlations identified using LINCS/CMAP connectivity analysis of proteome profiles from CNA amplified (AMP) or deleted (DEL) samples. The bar plots show, for each potential driver gene, the total number of statistically significant *trans* correlations (FDR < 0.05), along with the number of *trans*-correlated genes ("cmap overlap") that are also outliers (absolute value of z-score > 1.96) in the LINCS/CMAP knockdown profiles. Sample annotations enriched in the sample subsets representing each potential driver gene are summarized in Table S3.

G) Histograms of gene-wise mRNA to protein Pearson correlations of all (gray) and significant correlations (FDR < 0.01) (red).

H) Multi-omic landscape of key BRCA genes, TP53, PIK3CA, ERBB2, ESR1, PGR and GATA3. The heatmap has been generated with the CPTAC-BRCA2020 data viewer (see Data and Code Availability) and resembles Figure 1c in the retrospective BRCA study (Mertins et al., 2016).
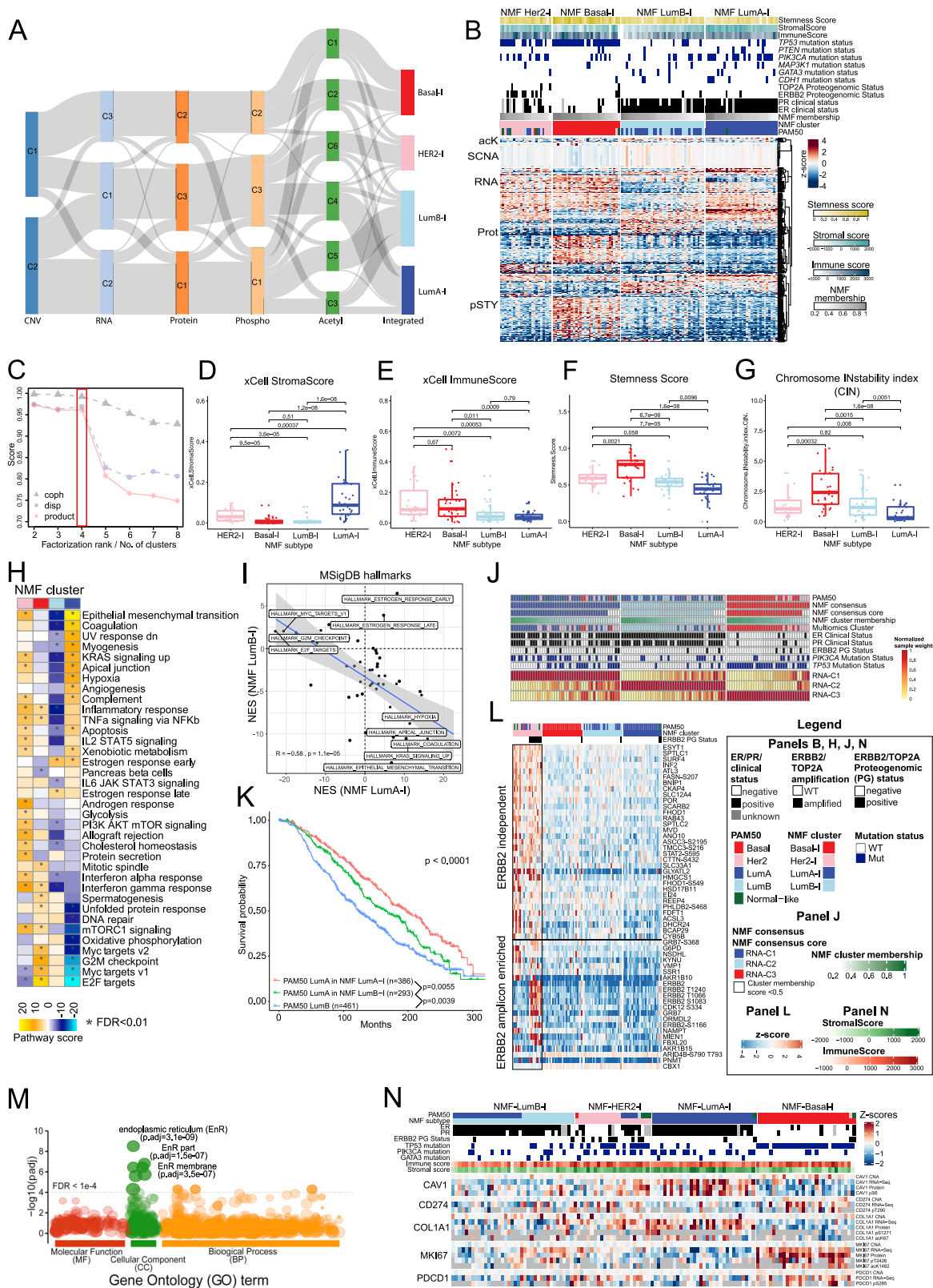
(legend on next page)

Figure S3. Molecular Subtyping, Related to Figure 1

A) Results of unsupervised non-negative matrix factorization (NMF) subtyping applied to individual data types (STAR Methods). The Sankey diagram depicts the flow of cluster assignments of all 122 tumors across data types.

B) Unsupervised multi-omics subtyping via NMF identified four molecular subtypes with distinct multi-omics expression patterns. The ability of NMF to assign weights to features that drive disease clustering (STAR Methods) allowed the identification of proteogenomic features (proteins, phosphorylation sites, acetylation sites, mRNA transcripts and SCNAs) that showed increased or decreased abundance levels specifically in each NMF cluster. The heatmap depicts the standardized abundances (z-scores) of 851 distinct features discriminating the molecular subtypes. acK = acetylated lysine sites; SCNA = somatic copy number alterations; RNA = gene-level transcripts, Prot = proteins; pSTY = phosphosites. Fisher's exact test was used to test for enrichment of clinical variables in core sets of tumors in each cluster. The NMF LumA-I cluster was enriched (Fisher's exact test) for hormone receptor positivity (estrogen receptor (ER), $p = 2.66 \times 10^{-5}$ and progesterone receptor (PR), $p = 3.55 \times 10^{-6}$) and wild-type $TP53$ ($p = 8.51 \times 10^{-5}$). The NMF Basal-I cluster was enriched for $TP53$ mutations ($p = 4.24 \times 10^{-10}$) and negative clinical hormone receptor status (ER, $p = 1.76 \times 10^{-10}$ and PR, $p = 1.02 \times 10^{-9}$). The NMF HER2-I was enriched for HER2-enriched PAM50 subtype samples ($p = 1.38 \times 10^{-8}$) and samples with centrally-confirmed, clinically positive ERBB2 status ($p = 3.97 \times 10^{-5}$). While NMF LumB-I and NMF LumA were only enriched for their respective PAM50 luminal subtype ($p = 1.51 \times 10^{-7}$ and $p = 3.95 \times 10^{-11}$, respectively) both luminal clusters were enriched for positive receptor status (ER [$p = 1.85 \times 10^{-6}$, $p = 1.85 \times 10^{-5}$] and PR [$p = 1.26 \times 10^{-3}$, $p = 3.55 \times 10^{-6}$]) as well as wild-type $TP53$ ($p = 9.88 \times 10^{-6}$, $p = 8.51 \times 10^{-5}$).

C) Graph showing cluster metrics (y axis) as a function of the number of clusters K (x axis). Cophenetic correlation (coph), dispersion of the consensus matrix (disp) and the product of both metrics are shown (product).

D) Boxplots depicting the distribution of stromal scores inferred by xCell between molecular subtypes. Boxes depict the interquartile range (IQR) of the scores with horizontal lines depicting the median. Whiskers extend to 1.5xIQR from Q1 (25th percentile) and Q3 (75th percentile), respectively. Points depict individual data points. P values were derived by pairwise Wilcoxon rank-sum tests.

E) Same layout as D), showing distribution of immune scores inferred by xCell.

F) Same layout as D), showing distribution of stemness scores.

G) Same layout as D), showing distribution of chromosomal instability index (CIN) scores.

H) Enrichment of Cancer Hallmark (Liberzon et al., 2015) gene sets in multi-omics subtypes detected by single sample gene set enrichment analysis (ssGSEA; Barbie et al., 2009) applied to feature weights determined by NMF to identify the dominant pathway signatures in each cluster. Normalized enrichment scores (NES) are shown. Asterisks indicate gene sets detected at FDR < 0.01.

I) Scatterplot comparing pathway (MSigDB Hallmark) enrichment scores between the LumA-I (x axis) and LumB-I (y axis) NMF subypes. Pathway scores were significantly anti-correlated (Pearson r = −0.58, $p = 1.1 \times 10^{-5}$).

J) NMF clustering applied to the RNA data in isolation. Heatmap depicting the relative contributions of each sample (x axis) to each cluster (y axis). This heatmap is a visualization of the (column-normalized) meta-feature matrix derived from decomposing the input matrix (STAR Methods), normalized per column by the maximum entry.

K) Kaplan-Meier plot showing survival outcome of PAM50 luminal A samples in the METABRIC database that were assigned by a Random Forest protein-based classifier to the NMF Lum-AI (red) or LumA/B subtypes (green) compared to PAM50 luminal B samples (blue) (similar to the Kaplan-Meier analysis in Figure 1C, but using a protein- rather than RNA-based Random Forest classifier). The significance of comparisons is noted. P values were derived from log-rank tests.

L) Heatmap depicting standardized abundances of features defining the NMF HER2-I cluster. Features (proteins and phosphosites) are designated to the right of the heatmap in descending rank according to the number of HER2-I samples that have standardized abundances for that feature above the average for the entire tumor cohort (z-score > 0). The horizontal line demarcates features up to the first occurence of an element from the $ERBB2$ amplicon itself (GRB7), which are therefore labeled "ERBB2-independent."
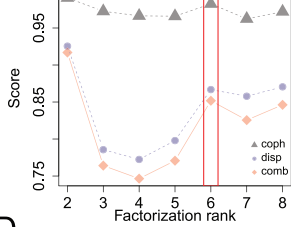
M) Overrepresentation analysis (hypergeometric test) of upregulated genes in the NMF HER2-I cluster that are independent of the $ERBB2$ amplicon (box in D). The Manhattan-like plot depicts log-transformed FDR-corrected p values (y axis) for enriched gene ontologies (GO) terms (Reimand et al., 2018). The dashed line indicates an FDR of 1×10^{-4}, which was used to define statistical significance for this analysis.

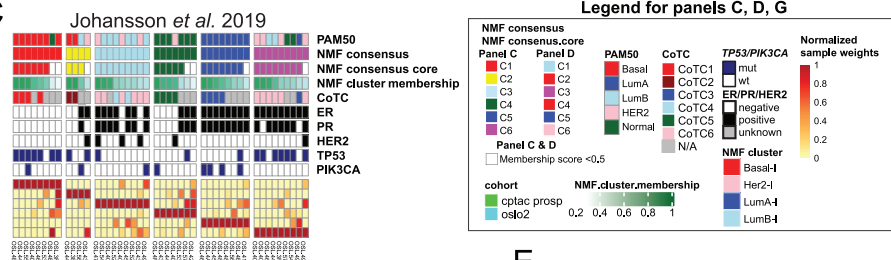N) Heatmap depicting multi-omics expression patterns of CD274, MKI67, PDCD1, CAV1 and COL1A1 (http://prot-shiny-vm.broadinstitute.org:3838/CPTAC-BRCA2020).

A

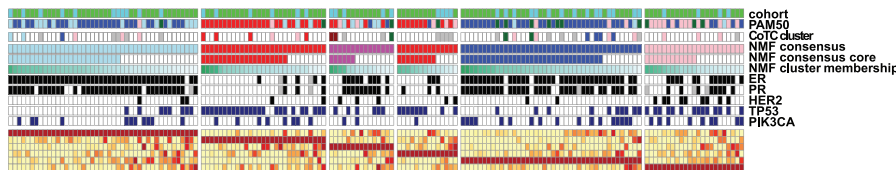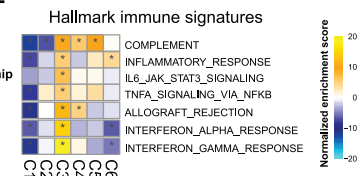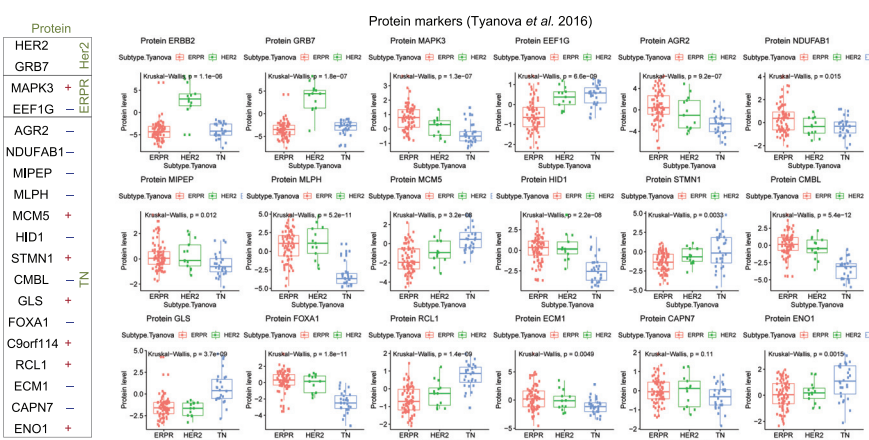|	This Study	Mertins et al., 2016	Johansson et al., 2019	Tyanova et al., 2016	Bouchal et al., 2019
# Tumors	122 fresh frozen tumors	77 fresh frozen tumors	45 fresh frozen tumors	40 FFPE	96
Quantitation	TMT-10	iTRAQ-4	TMT-10	Super-SILAC	LFQ (SWATH-MS)
Fractionation	bRPLC	bRPLC	HiRIEF	SAX	
Type (# Fractions)	Proteome (24) Phosphoproteome (12) Acetylproteome (6)	Proteome (24) Phosphoproteome (12)	Proteome (72)	Proteome (6)	Proteome (One shot and matched to a fractionated (20 fractions) spectral library)
Gradient length	110 / 110 / 160 min	110 / 110 min	60-90 minutes	200 min	120 min
Instrument	Orbitrap Fusion Lumos	Orbitrap Q Exactive	Orbitrap Q Exactive	Orbitrap Q Exactive	TripleTOF 5600+
Proteomic depth	10,107 proteins (7,767 full quant)	10,915 proteins (7,472 full quant)	13,997 proteins (9,995 full quant)	10,138 proteins (2,588 full quant)	2,842 proteins (full quant)
	38,968 phosphosites (8,778 full quant)	33,122 phosphosites (2,790 full quant)	---	---	---
	9,869 acetylsites (465 full quant)	---	---	---	---
Subtype composition	HER2-E: 14 (11.5%) Basal-like: 29 (23.8%) Luminal A: 57 (46.7%) Luminal B: 17 (13.9%) Normal-like: 5 (4.1 %)	HER2-E: 15.6% Basal-like 23.4% Luminal A: 29.9% Luminal B: 31.2% Normal-like: -	HER2-E: 9 (20%) Basal-like: 9 (20%) Luminal A: 9 (20%) Luminal B: 9 (20%) Normal-like: 9 (20%)	ER+ / PR+: 14 (35%) Her2+: 15 (37.5%) TN: 11 (27.5%)	HER2-E: 8 (8.3%) TN: 16 (16.7%) LumA: 48 (50%) LumB: 16 (16.7%) LumB-HER2+: 8 (8.3%)
Unsupervised proteomic / multiomic subtypes	Multi-omics subtypes: NMF-LumA-I: 32 (26.2%) NMF-LumB-I: 37 (30.3%) NMF-Her2-I: 23 (18.9%) NMF-Basal-I: 30 (24.6%)	Basal-enriched: 19 (23.7%) Luminal-enriched: 31 (38.8%) Stromal-enriched: 30 (37.5%)	CoTC1 (basal-like): 4 CoTC2 (basal-like): 2 CoTC3 (LumA): 9 CoTC4 (LumB): 2 CoTC5 (Normal-like): 4 CoTC6 (LumB+HER2): 12	Partial co-clustering of the same classical subtype (Figure 1d in Tyanova et al 2016)	Analysis of pooled samples from five subtypes; High similarity between ER+ and HER2-; LumA and LumB; ER- and HER2+ and TN (Figure 1A in Bouchal et al 2019)

B

C
Johansson et al. 2019

Legend for panels C, D, G

D

E
Hallmark immune signatures

F
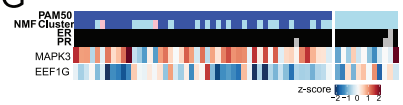Protein markers (Tyanova et al. 2016)
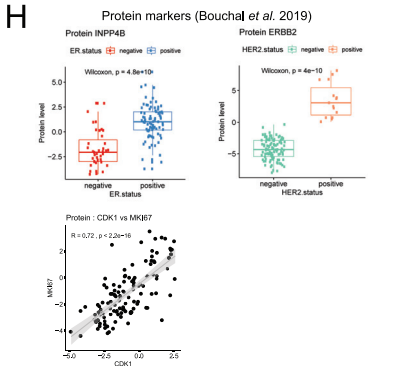
G

H
Protein markers (Bouchal et al. 2019)

(legend on next page)

Figure S4. Comparison with Previous BRCA Proteomics and PG Studies, Related to Figures 1 and S3

A) Summary of this and prior studies utilizing proteomics or proteogenomics to profile and cluster breast tumors. Shown are sample size analyzed, mass spectrometry platform and proteomic workflow, proteomic depth achieved, subtype composition and unsupervised clustering results.

B) NMF clustering applied to the proteome dataset from Johansson et al. (2019). Their study of 45 breast tumors with balanced representation of all five PAM50 subtypes described six stable proteomic clusters derived from the subset of 29 tumors that had Pearson correlation $\geq$ 0.5 relative to at least one other tumor (Johansson et al., 2019). In addition to an alignment to the PAM50 subtypes, the authors reported a further subdivision of poor-prognosis basal-like and luminal B tumors driven by immune component infiltration. Application of the NMF pipeline to their complete proteomic data likewise identified six proteome clusters. Graph shows cluster metrics (y axis) as a function of the number of clusters (K, x axis). Cophenetic correlation (coph), dispersion of the consensus matrix (disp) and the product of both metrics are shown (product).

C) Heatmap depicting the relative contributions of each sample (x axis) to each cluster (y axis) resulting from application of the NMF pipeline to Johansson et al. (2019). This heatmap is a visualization of the (column-normalized) meta-feature matrix derived from decomposing the input matrix (STAR Methods), normalized per column by the maximum-valued entry. High concordance was observed between consensus core tumor clusters (CoTC) reported in Johansson et al. (2019) and NMF-based subtyping.

D) NMF clustering applied to an integrated proteomic dataset derived from Johansson et al. and the present study (STAR Methods). The comparable MS workflows utilized in the studies together with the similar cluster structures revealed by the different analytical approaches enabled such integration. Only four PAM50 LumA samples from Johansson et al. (2019) were assigned to the NMF LumA-I cluster; of the remaining five assigned to the NMF LumB-I cluster, four had been excluded from their clustering exercise because of low correlation to other samples. These data support our contention that NMF clustering identifies a smaller, more homogeneous LumA-I group by reassigning some PAM50 LumA samples into the LumB-I group.
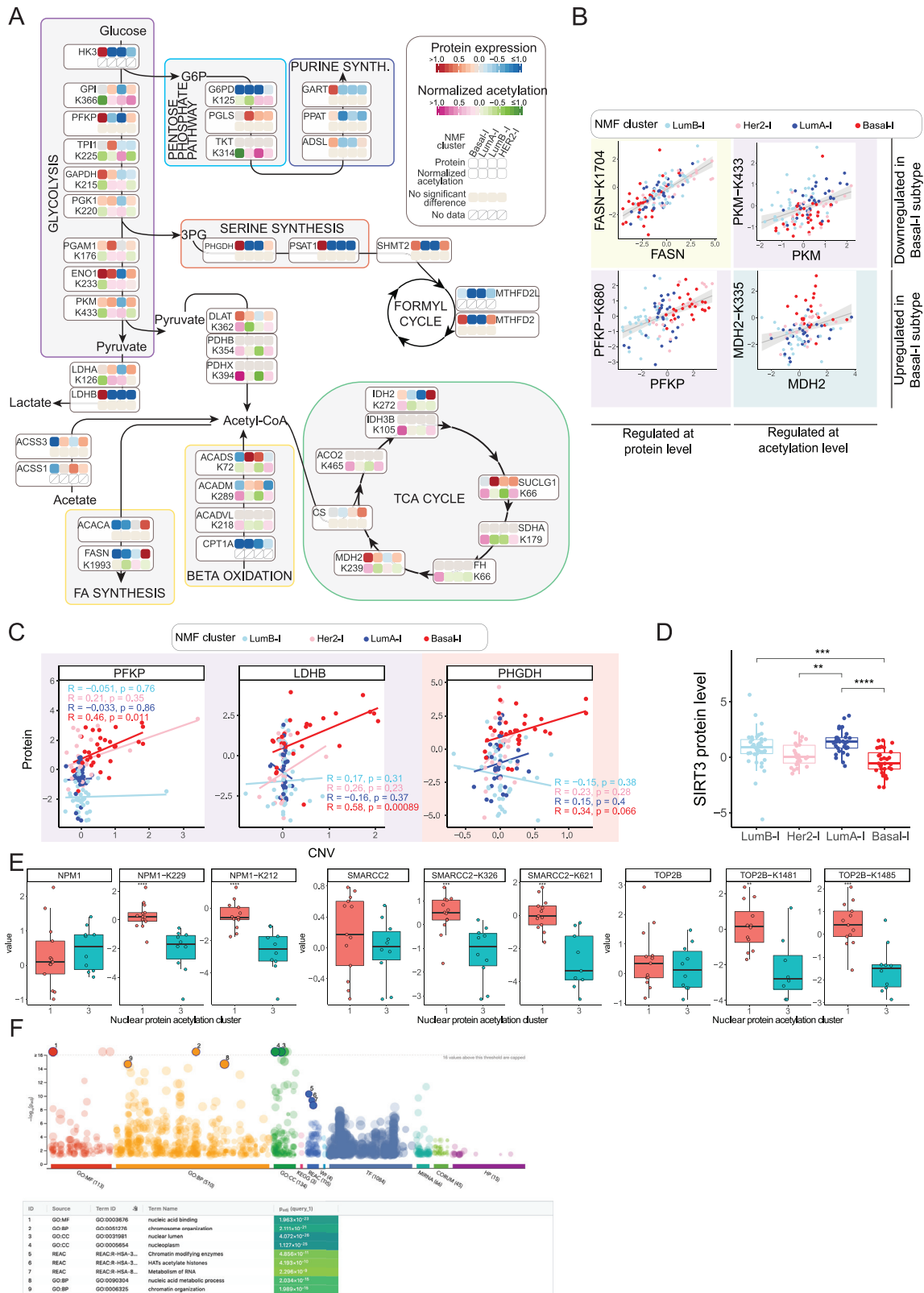
E) Normalized enrichment scores (NES) of immune-related Hallmark signatures based on the feature weights driving the clustering shown in D) (STAR Methods). The two samples that formed the Johansson et al. CoTC2 ("basal immune") cluster joined a cluster in the combined dataset that was much more heterogeneous in PAM50 composition but had immune-related Hallmark gene signatures as a unifying characteristic. We therefore suggest that an active immune microenvironment rather than the basal subtype assignment was critical to the definition of the CoTC2 cluster.

F) Tyanova et al. (2016) identified a 19-protein predictive classifier for hormone receptor status-based subtypes (ERPR = ER+/PR+; HER2 = HER2+; TN = ER-/PR-/HER2-) by applying support vector machine (SVM)-based classification to super-SILAC proteomic profiles of 40 total ER/PR+, HER2 and triple negative breast tumors. The left box shows all 19 proteins grouped by subtype and annotated by their predictive abundance relative to the other subtypes: increased (+) or decreased (-) expression. The boxplots depict the abundance of 18 of these proteins quantified in our study across ERPR, HER2 and TN subtypes according to our clinical (ER, PR) and proteogenomic-based (HER2) assessment of receptor status (C9orf114 was not quantified in our analysis, which excluded 5 samples for which receptor status annotation was incomplete). A negative HER2 status was required for tumors classified as ERPR. Seventeen out of 18 measured markers maintained their subtype associations in our dataset (Kruskal-Wallis test). Only CAPN7 did not replicate, showing a trend as a negative marker of TNBC that did not reach statistical significance. Boxes depict the interquartile range (IQR) of the data with horizontal lines depicting the median. Whiskers extend to 1.5xIQR from Q1 (25th percentile) and Q3 (75th percentile), respectively. Points depict individual data points.

G) Heatmap depicting standardized protein abundances of MAPK3 and EEF1G across all tumors classified as ERPR in our dataset. These positive and negative ER/PR markers captured PAM50 LumA but interestingly not LumB samples (a distinction not made in the Tyanova et al., 2016 analysis).

H) Bouchal et al. (2019) reported three "proteotypic" proteins associated with ER (INPP4B) and HER2 (ERBB2) receptor status and with tumor grade (CDK1) in a study using a label-free, data-independent MS acquisition method to examine 96 breast cancer specimens. The boxplots illustrate the protein abundance of INPP4B between ER+/− and ERBB2 between HER2+/− in our dataset, reproducing their results (Boxplots are defined in panel F). P values were derived from a Wilcoxon rank-sum test. Due to incomplete grade annotation in our dataset we did not test their association of CDK1 protein expression and grade; however, CDK1 was significantly correlated with MKi67, (Pearson R = 0.72), consistent with the correlation of R = 0.79 that they reported as a confirmatory analysis in the TCGA breast cancer dataset (scatterplot).

A

B

C

D

E

F

ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0003676	nucleic acid binding	1.963×10^{-23}
2	GO:BP	GO:0051276	chromosome organization	2.111×10^{-21}
3	GO:CC	GO:0031981	nuclear lumen	4.072×10^{-19}
4	GO:CC	GO:0005654	nucleoplasm	1.127×10^{-18}
5	REAC	REAC:R-HSA-3...	Chromatin modifying enzymes	4.856×10^{-17}
6	REAC	REAC:R-HSA-3...	HATs acetylate histones	4.193×10^{-15}
7	REAC	REAC:R-HSA-B...	Metabolism of RNA	2.296×10^{-9}
8	GO:BP	GO:0090304	nucleic acid metabolic process	2.034×10^{-16}
9	GO:BP	GO:0006325	chromatin organization	1.989×10^{-16}

Figure S5. PG Metabolic Profiling, Related to Figure 2

A) Pathway schematic similar to Figure 2B but showing differentially expressed metabolic proteins and normalized acetylation sites (Wilcoxon test FDR p value < 0.05) mapped onto key metabolic pathways across all four NMF subtypes.
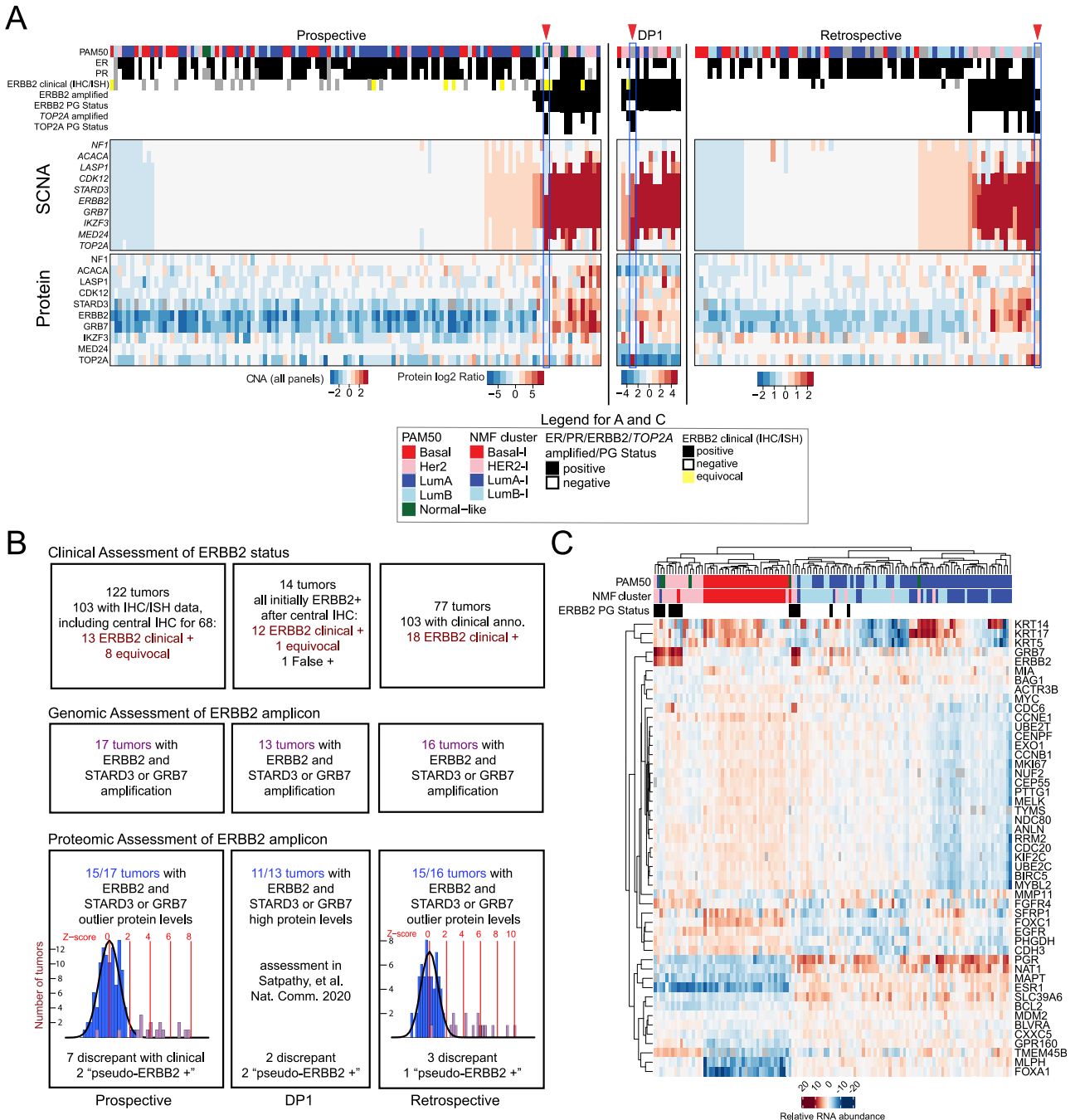
B) Scatterplots showing examples of proteins regulated at the protein or at the acetylation level across subtypes. Background color corresponds to pathway colors in Figure 2C.

C) Scatterplots showing correlation between copy number variation and protein expression of genes involved in metabolism. Spearman's Rho value is shown. Background color corresponds to pathway colors in Figure 2C.

D) Boxplot showing SIRT3 protein expression in four NMF subtypes. Wilcoxon test FDR p values are shown: **** $\leq$ 0.0001, *** $\leq$ 0.001, ** $\leq$ 0.01. Boxplots show 1.5x the interquartile range for each group, centered on the median.

E) Further examples of nuclear proteins that were differentially acetylated between subtypes 1 and 3 (see Figure 2E), but were not differential at the protein level. Boxplots show 1.5x the interquartile range for each group, centered on the median.

F) gProfiler pathway and GO overrepresentation analysis of proteins from the 'Acetylation up in cluster 1' group in Figure 2G. Proteins were enriched for nucleoplasm, chromatin organization, HATs acetylate histones, and Metabolism of RNA pathways.
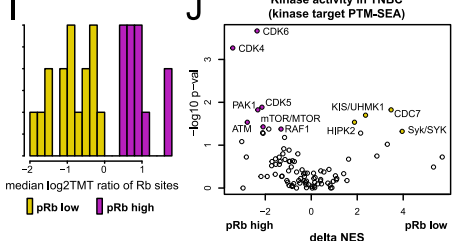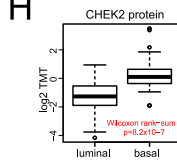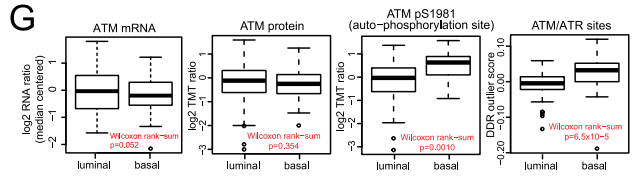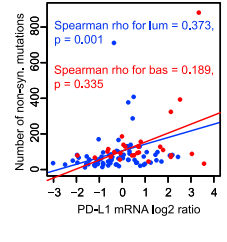
A

Prospective | DP1 | Retrospective

PAM50
ER
PR
ERBB2 clinical (IHC/ISH)
ERBB2 amplified
ERBB2 PG Status
TOP2A amplified
TOP2A PG Status

SCNA
NF1
ACACA
LASP1
CDK12
STARD3
ERBB2
GRB7
IKZF3
MED24
TOP2A

Protein
NF1
ACACA
LASP1
CDK12
STARD3
ERBB2
GRB7
IKZF3
MED24
TOP2A

CNA (all panels) −2 0 2
Protein log2 Ratio −5 0 5
−4 −2 0 2 4
−2 −1 0 1 2

Legend for A and C

PAM50	NMF cluster	ER/PR/ERBB2/*TOP2A* amplified/PG Status	ERBB2 clinical (IHC/ISH)
Basal	Basal-I	positive	positive
Her2	HER2-I	negative	negative
LumA	LumA-I		equivocal
LumB	LumB-I		
Normal−like			

B

Clinical Assessment of ERBB2 status

122 tumors 103 with IHC/ISH data, including central IHC for 68: 13 ERBB2 clinical + 8 equivocal	14 tumors all initially ERBB2+ after central IHC: 12 ERBB2 clinical + 1 equivocal 1 False +	77 tumors 103 with clinical anno. 18 ERBB2 clinical +

Genomic Assessment of ERBB2 amplicon

17 tumors with ERBB2 and STARD3 or GRB7 amplification	13 tumors with ERBB2 and STARD3 or GRB7 amplification	16 tumors with ERBB2 and STARD3 or GRB7 amplification

Proteomic Assessment of ERBB2 amplicon

15/17 tumors with ERBB2 and STARD3 or GRB7 outlier protein levels	11/13 tumors with ERBB2 and STARD3 or GRB7 high protein levels assessment in Satpathy, et al. Nat. Comm. 2020	15/16 tumors with ERBB2 and STARD3 or GRB7 outlier protein levels
7 discrepant with clinical 2 "pseudo-ERBB2 +"	2 discrepant 2 "pseudo-ERBB2 +"	3 discrepant 1 "pseudo-ERBB2 +"
Prospective	DP1	Retrospective

C

PAM50
NMF cluster
ERBB2 PG Status

KRT14
KRT17
KRT5
GRB7
ERBB2
MIA
BAG1
ACTR3B
MYC
CDC6
CCNE1
UBE2T
CENPF
EXO1
CCNB1
MKI67
NUF2
CEP55
PTTG1
MELK
TYMS
NDC80
ANLN
RRM2
CDC20
KIF2C
UBE2C
BIRC5
MYBL2
MMP11
FGFR4
SFRP1
FOXC1
EGFR
PHGDH
CDH3
PGR
NAT1
MAPT
ESR1
SLC39A6
BCL2
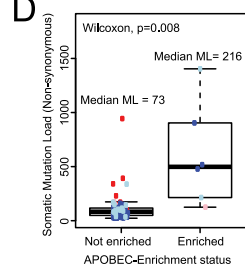MDM2
BLVRA
CXXC5
GPR160
TMEM45B
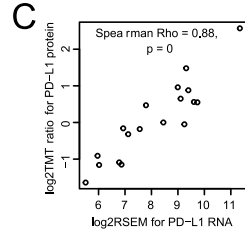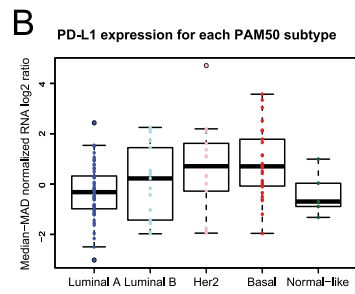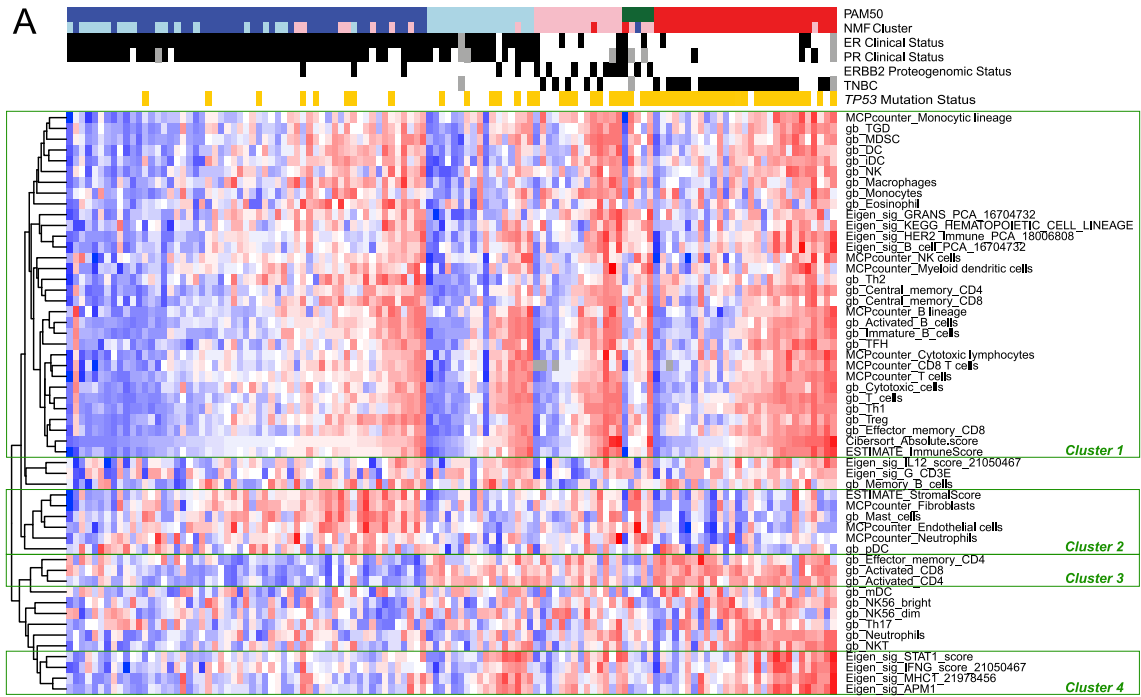MLPH
FOXA1

Relative RNA abundance

Figure S6. PG Classification of HER2, Related to Figure 3
A) Proteogenomic analysis of the *ERBB2* amplicon for all samples from each dataset shown in Figure 3A.
B) Assessment and frequencies of clinical ERBB2 positive/equivocal samples, *ERBB2* gene-amplified, and ERBB2 proteogenomically (PG) positive samples in each dataset shown in Figure 3A. Clinical assessment was refined after centralized IHC (see STAR Methods), and discrepant cases represent those for which clinical status was ERBB2 positive or equivocal, but PG status was negative. "Pseudo-ERBB2 +" cases are those for which amplicon genes were amplified without a corresponding increase in protein expression.
C) Heatmap illustrating the RNA abundance levels of 49 PAM50 genes detected in this dataset. Rows and columns are hierarchically clustered.

A

B PD-L1 expression for each PAM50 subtype

C

D

E

F Prospective   Retrospective

G

H

I

J Kinase activity in TNBC (kinase target PTM-SEA)
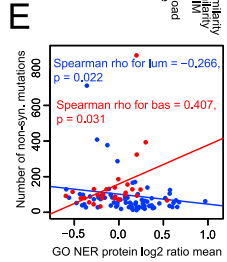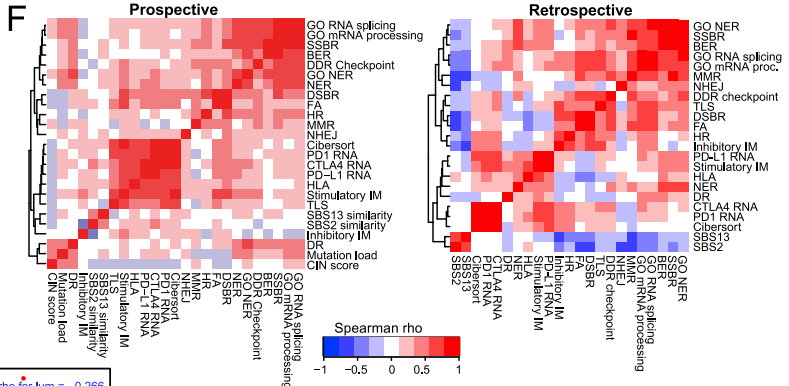
Figure S7. Immune Landscape of BRCA and Phospho-Rb-Dependent Kinase Activity in TNBC, Related to Figures 4, 5, and 6

A) Heatmap of scores for immune and stromal cell types inferred from RNA-seq data by MCP-Counter (labeled MCPcounter; Becht et al., 2016a, 2016b) and by the immunophenotyping approach described by Angelova et al. (2015) (labeled gb), as well as for protein-based Eigen signatures from Thorsson et al. (2019) (labeled Eigen_sig). Table S6 includes these profile scores, as well as scores for cell types inferred by CIBERSORT and xCell, which showed similar patterns but were excluded from this figure for the sake of clarity and because they had a high prevalence of 0 values. Hierarchical clustering by Euclidean distance across all samples separated these signatures into four clusters: Cluster 1: Immune signatures that track with Cibersort absolute score in all PAM50 subtypes. Cluster 2: Immune signatures that track with Cibersort in PAM50 LumA tumors but show an overall reduction in levels in Basal and LumB tumors, including stromal (ESTIMATE), fibroblast (MCPcounter), mast cell (Angelova et al., 2015), endothelial cell (MCPcounter), and neutrophil (MCPcounter) signatures (p = 2.17×10^{-4}, 1.31×10^{-4}, 6.35×10^{-6}, 9.63×10^{-7}, 2.51×10^{-3}, respectively for cell type, Wilcoxon rank sum tests comparing LumA to Basal + LumB). Cluster 3: Effector memory and activated CD4+ and activated CD8+ signatures (Angelova et al., 2015) that were lower in the PAM50 LumA subtype than in all other subtypes (p = 5.79×10^{-7}, 9.29×10^{-15}, and 4.86×10^{-14}, respectively, Wilcoxon rank sum tests comparing LumA to all other tumors). Cluster 4: Interferon gamma (IFNG), STAT1, MHC1, and antigen presentation machinery (APM1) protein signatures (Thorsson et al., 2019) that tracked with Cibersort immune score in all PAM50 subtypes except LumA, where they were lower than in Basal (p = 2.08×10^{-3}, 6.67×10^{-3}, 6.81×10^{-5}, and 2.75×10^{-5}, respectively, Wilcoxon rank sum tests comparing LumA and Basal).

B) Boxplot of PD-L1 mRNA expression in each PAM50 subtype. Boxplots show 1.5x the interquartile range for each group, centered on the median.

C) Scatterplot showing that PD-L1 mRNA was well correlated (Spearman) with PD-L1 protein in the DP1 study (Satpathy et al., 2020).

D) Boxplot comparing mutational load between APOBEC-enriched and -unenriched samples. The boxplot shows 1.5x the interquartile range for each group, centered on the median. P value is from the Wilcoxon rank-sum test.

E) Scatterplots showing that the GO BP nucleotide excision repair pathway (GO:0006289) is negatively correlated with number of non-synonymous mutations (top), which in turn is positively correlated with PD-L1 mRNA (bottom), in luminal but not basal samples. For clarity, one hyper-mutated basal sample was excluded from these plots. Spearman's rank correlation coefficients (rho) and p values are shown.

F) Heatmaps for pairwise Spearman rank correlations between the features shown in Figure 5E within samples from the PAM50 basal subtype in the prospective (left heatmap) and retrospective (right heatmap) datasets. NER: nucleotide excision repair, BER: base excision repair, MMR: mismatch repair, DR: direct repair, TLS: translesion synthesis, NHEJ: non-homologous end joining, FA: Fanconi anemia, HR: homologous recombination, DDR: dna damage response (primarily checkpoint proteins), SSBR: single strand break repair, DSBR: double strand break repair.

G) Multi-omics comparisons of the ATM DNA damage checkpoint kinase in HR+/ERBB2- samples and TNBC samples. ATM mRNA (1st panel) and protein (2nd panel) levels were unchanged (Wilcoxon rank sum tests: p = 0.052 and p = 0.354, respectively) while the levels of a known auto-phosphorylation site (3rd panel) and the ATM/ATR target site outlier scores (4th panel) were significantly lower in HR+/ERBB2- than in TNBC tumors (Wilcoxon rank sum tests: p = 0.010 and p = 6.5×10^{-5}, respectively). Boxplots show 1.5x the interquartile range for each group, centered on the median.

H) Chk2 (CHEK2) protein levels were significantly lower in HR+/ERBB2-luminal than in TNBC-basal tumors (Wilcoxon rank sum tests: p = 8.2×10^{-7}). The boxplot shows 1.5x the interquartile range for each group, centered on the median.

I) Histogram showing distribution of Rb phosphoprotein (mean of add Rb sites) levels in TNBC samples. Samples were classified as pRb low (yellow) and pRb (purple) high for the analysis in J.

J) CDK4 and CDK6 kinase activity scores inferred from phosphosite data by PTM-SEA were significantly lower in TNBC samples with low Rb phosphorylation than in TNBC samples with high Rb phosphorylation (p = 3.5×10^{-5} for CDK4 activity; p = 7.3×10^{-7} for CDK6 activity). p values are from Wilcoxon rank sum tests comparing the two groups shown in I. Volcano plot shows the -log 10 p values (y axis) versus differences for kinase activity scores inferred by PTM-SEA (kinase target phosphosite sets; x axis). Kinases with p < 0.05 are labeled and colored by higher activity in Rb high (purple) or in Rb low (yellow) samples. delta NES: difference in median activity score for pRb low and pRb high samples.