**Supplementary data**

**Table S1. Clinical information and CMS classification profile of publicly collected TNBC datasets.**

**Table S2. Gene set list, expression mean and SVM information gain for CMS classification**

**Table S3. Further survival analysis result of TNBC subtypes**

**Table S4. DEG set comparing primary with metastasis and metastasis prediction signature gene set for random-forest machine-learning**

**Table S5. Two GSEA results using DEG of primary and metastasis, and metastasis prediction signature gene set**

**Table S6. Clinical information and CMS classification profile of ROADMAP project patients**

**Table S7. Pathways related to cisplatin resistance genes derived from ROADMAP expression profiles**

**Table S8. Differentially expressed genes involved in cisplatin resistance inferred from ROADMAP gene expression profile**

**Figure S1. Gene expression analysis and subtype classification**

**a** PCA plot of gene expression before and after normalization. Colors were assigned by GEO series ID. **b** Silhouette plot to assess non-negative matrix factorization clustering. **c** Bar plot representing classification performance. **d** Mosaic plot comparing Lehmann and Burstein classification status with our CMS. **e** Wilcoxon test of each subtype signature score with CMS; Lehmann and Burstein. We selected MSL, LAR, IM, and BL1 subtypes to show the best signature score from 6 subtypes using Lehmann’s method.

**Figure S2. Immune cell activation and prognostic status for each CMS**

**a** Heatmap of 22 leukocyte deconvolution from bulk tumors. **b** Disease-free and overall survival plots based on *CD8* T cell activation. **c** Gene expression boxplot and **d** Plotof Tukey’s range test of known immunotherapy markers *CD52* and *CTLA4*.

**Figure S3. Gene expression status of tumor-associated stromal cell makers**

*P-values* were calculated by ANOVA test.

**Figure S4. Key pathway status of SL subtype according to *BRCA1/2* germline mutation of TCGA dataset**

Boxplot of pathway activity scores with *BRCA1/2* germline mutation of TNBC samples in TCGA BRCA.

**Figure S5. Classification with metastatic prognosis genes and survival plot in primary samples**

**a** Overview process to identify metastatic prognosis genes. **b** AUC curves representing classification performance with the number of genes according to P-value threshold. **c** Significantly enriched pathways derived from the metastasis prediction markers. **d** Metastasis free survival plots of metastatic prognosis markers *SHC1, JAG2* and *ITGA1* according to gene expression.

**Figure S6. Characteristics of CMS-interferential drug responses and drug-pathway associations in our data set and biobank**

**a** Heatmaps of subtype-related DR scores and drug-pathway correlations. Drug responses of our meta-data to show difference among CMSs resemble with biobank. Patients and organoids of heatmap were ordered by CMS, and drugs were arranged by ANOVA test p-values and DR score status for each CMS (left). Correlations between pathway activities and DR scores divided pathways like CMS (right). Colors of correlation heatmap indicate meta-data correlation values and the circle size is biobank. Four yellow boxes for each left heatmap represent drug clusters assigned to our CMSs. Red font DR scores represents resistance, while blue indicates sensitive. **c** Clinical outcome summary of response signatures. Subtype-related DR scores confirmed in both our meta-data and biobank and partially implied on clinical outcomes MFS, CI, and HR.

**Figure S7. Correlation with cisplatin signature and CIN**

The correlation plots between cisplatin resistance and response DR scores, respectively. *P*-values and correlation coefficients were calculated by Spearman correlation test.

**Figure S8. Subtype classification and prognostic markers derived from ROADMAP data**

**a** Enrichment status of representative signature in ROADMAP expression profile. **b** Boxplot of key pathway activity scores to show differences among CMSs. **c** Progression-free survival plot and **d** CI and HR forest plots based on gene expression in ROADMAP data and aggregated data. Each red vertical line represents the average score with all markers, CI, and HR, respectively.