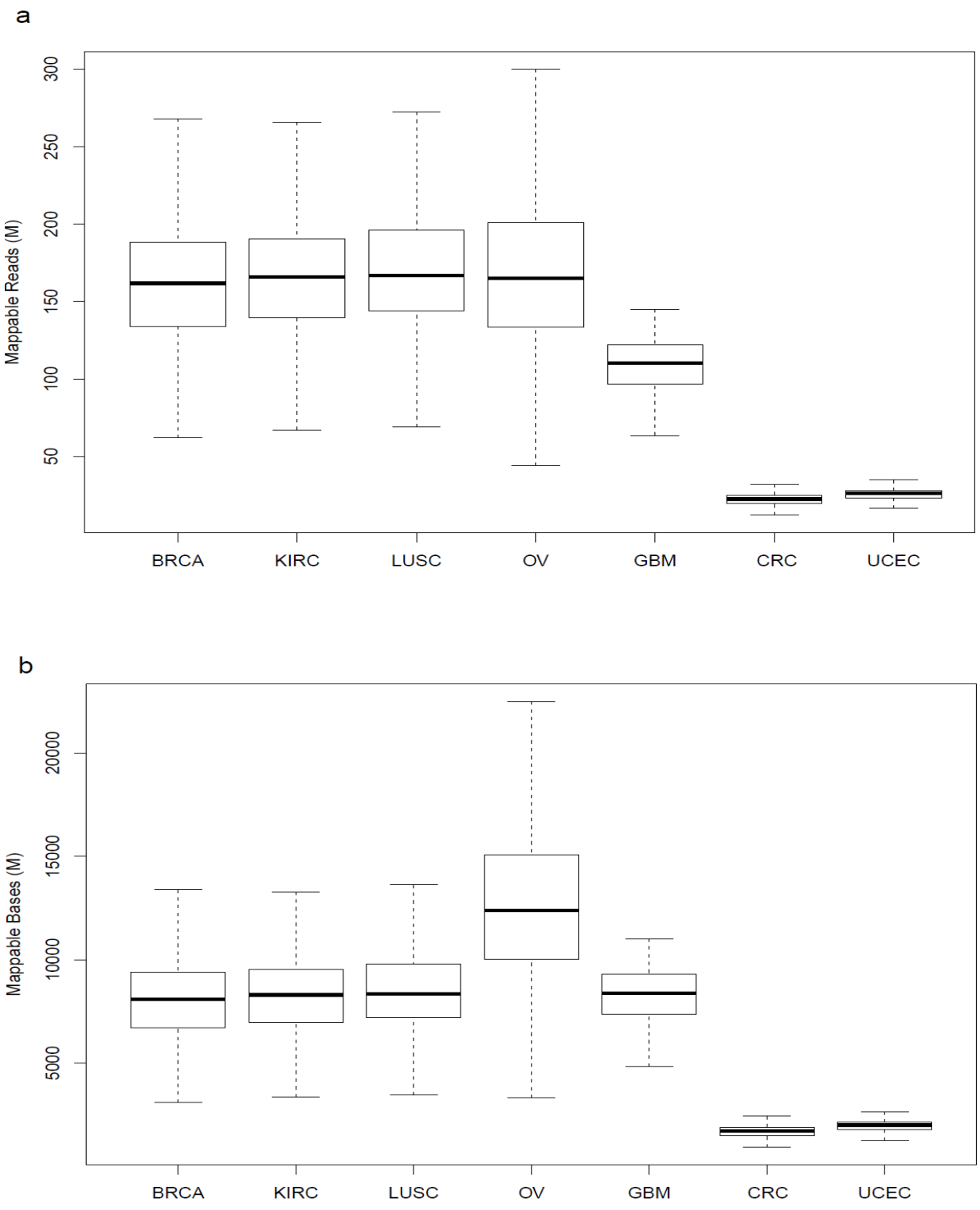
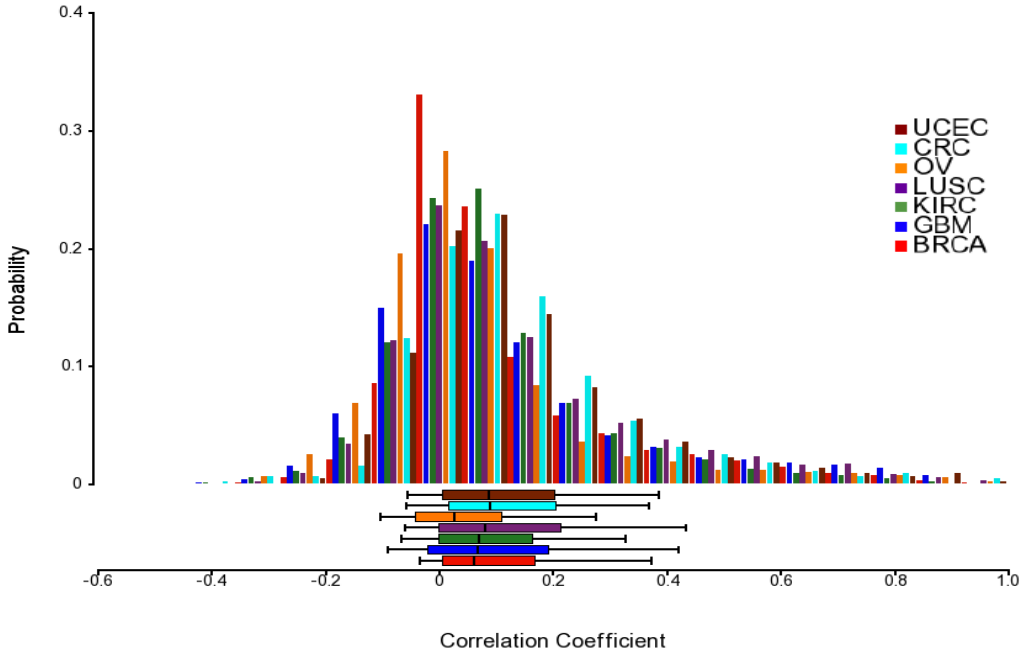


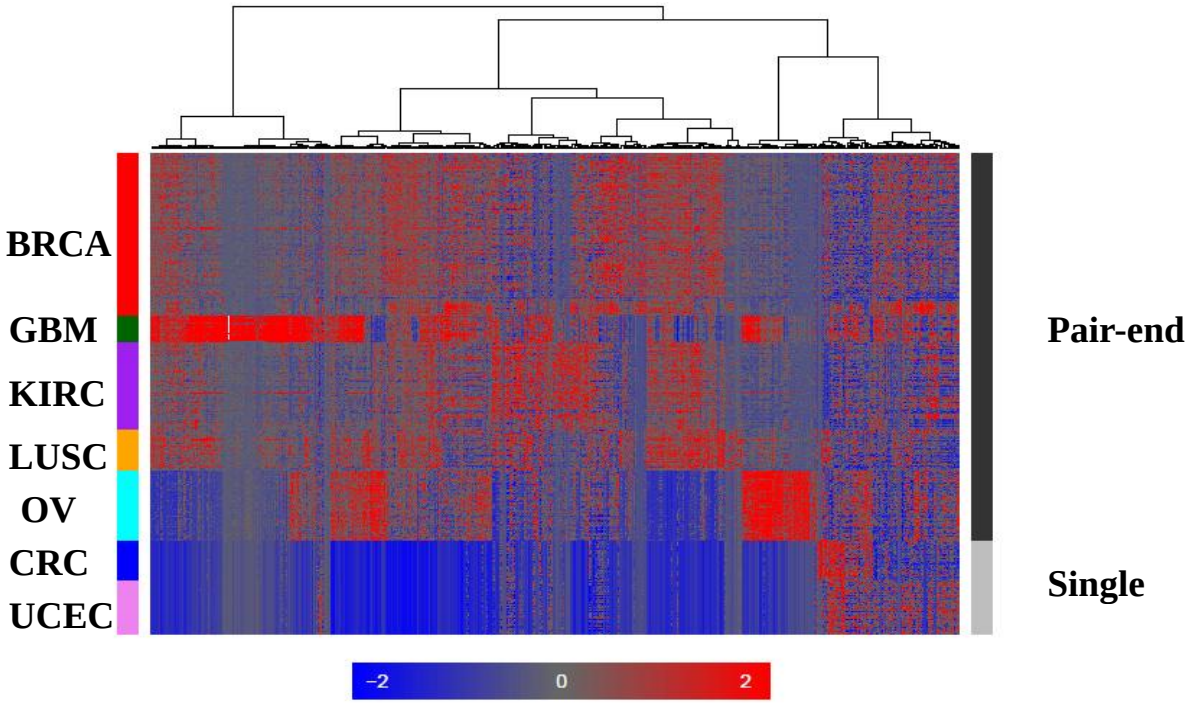

**Supplementary Figure 1. Distribution of mappable reads (million) and mappable (million) bases across different cancer types**

BRCA (*n* = 942), KIRC (*n* = 515), LUSC (*n* = 237), OV (*n* = 412), GBM (*n* = 154), CRC (*n* = 228), and UCEC (*n* = 320). The boxes show the median ± 1 quartile, with whiskers extending to the most extreme data point within 1.5 interquartile range from the box boundaries.

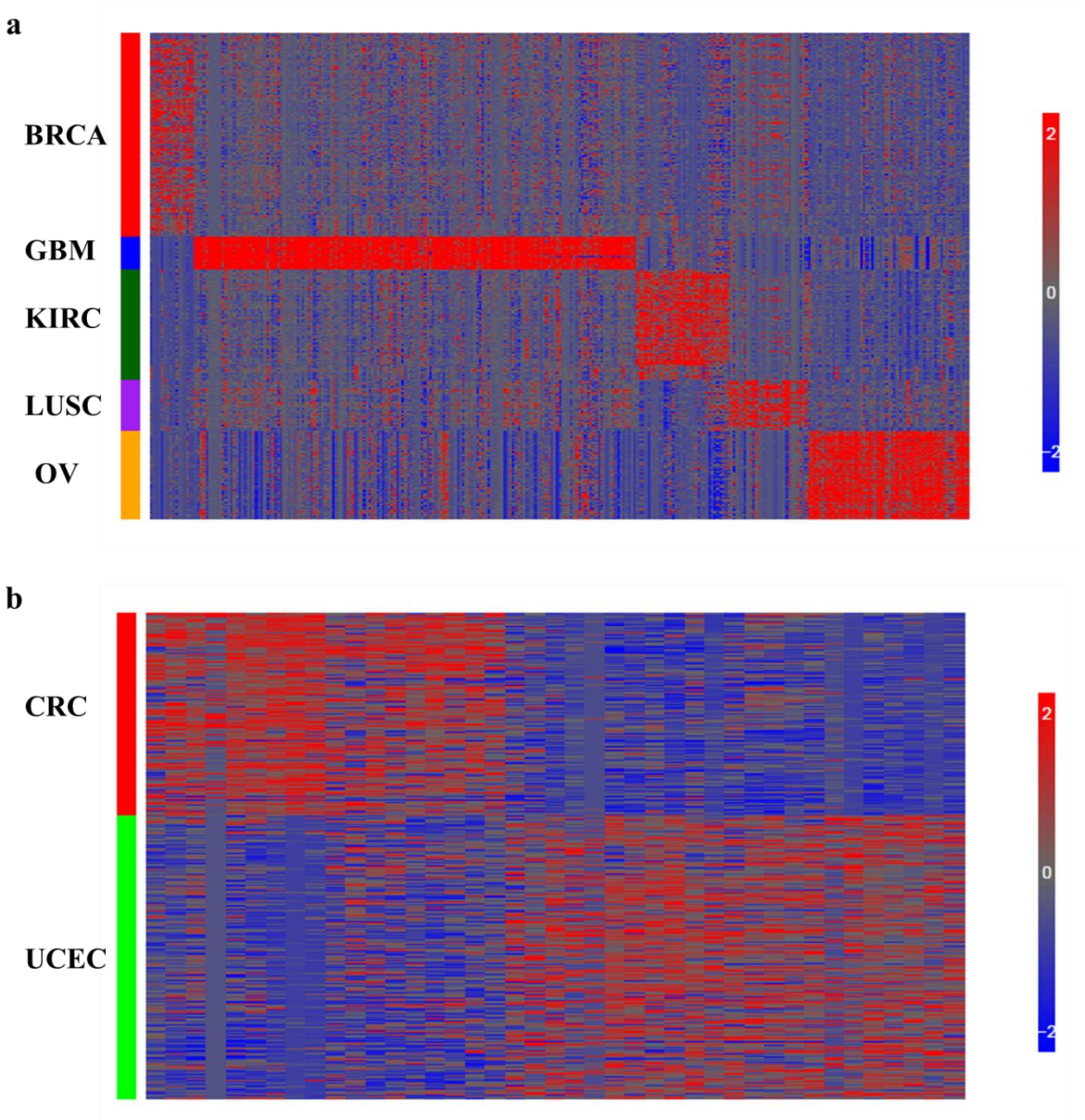



**Supplementary Figure 2. Correlations between pseudogenes and their WT cognate genes in different cancer types**

The boxes show the median ± 1 quartile, with whiskers extending to the most extreme data point within 1.5 interquartile range from the box boundaries.

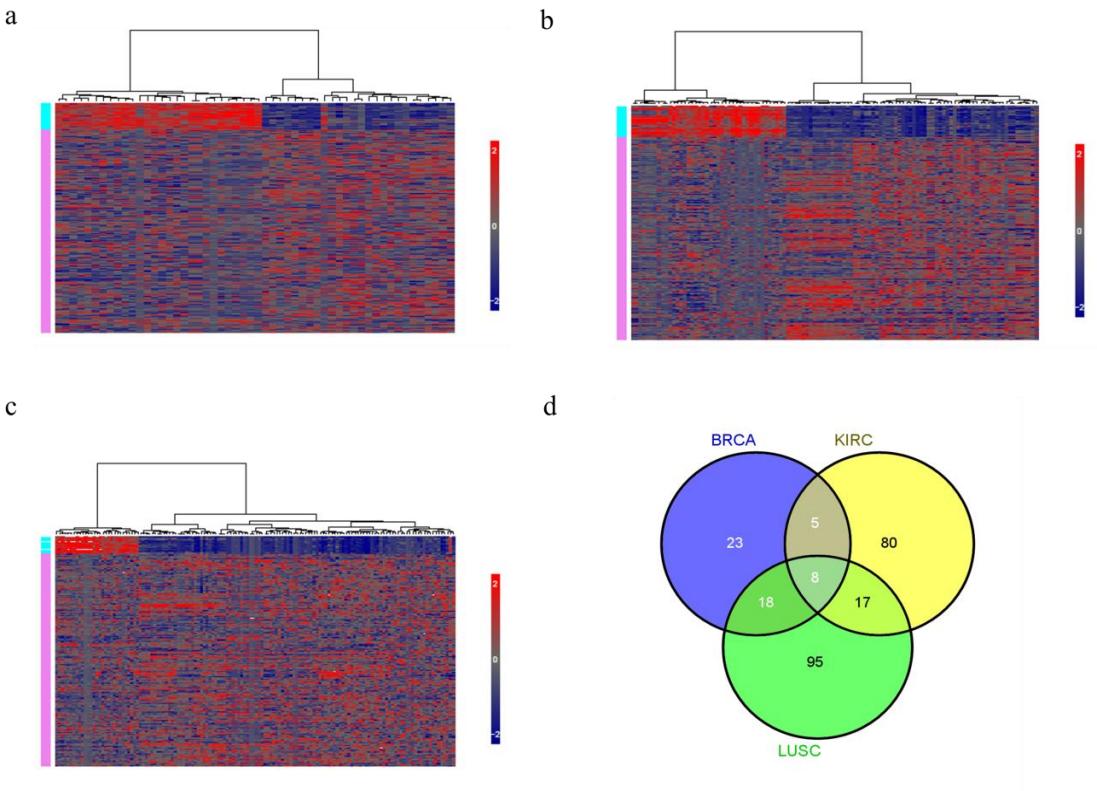



**Supplementary Figure 3. Global pattern of pseudogenes expression across different cancer types** Unsupervised clustering based on all pseudogenes surveyed. BRCA (*n* = 837), KIRC (*n* = 448), LUSC (*n* = 220), OV (*n* = 412), GBM (*n* = 154), CRC (*n* = 228), and UCEC (*n* = 316).



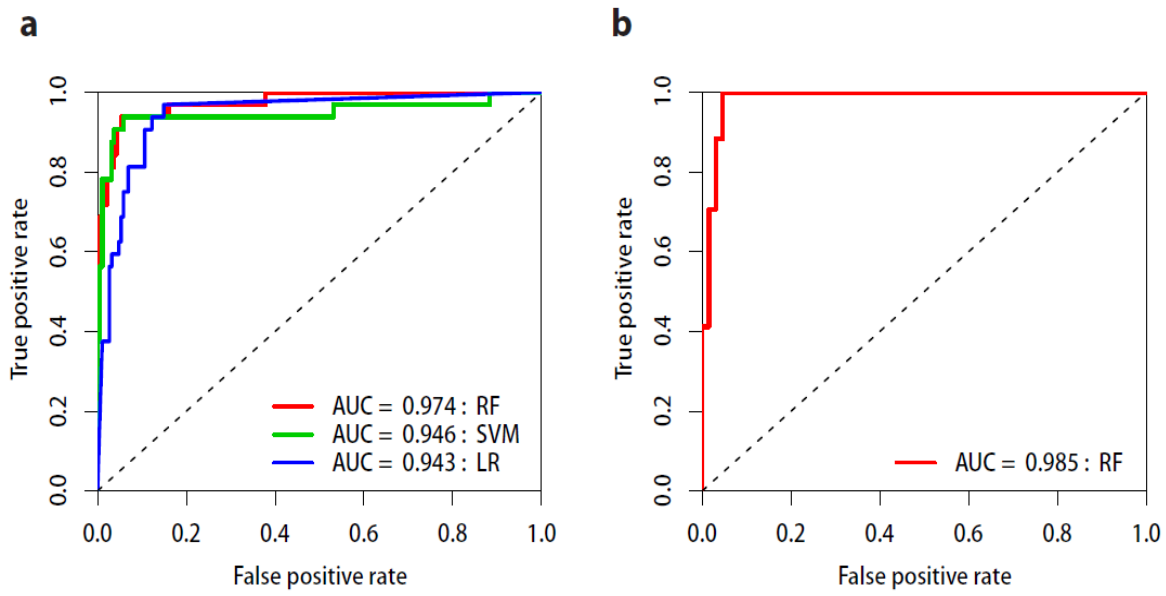

**Supplementary Figure 4. Tumor-lineage-specific pseudogenes expression across different cancer types**

(a) Pair-end sequencing for BRCA (*n* = 837), KIRC (*n* = 448), LUSC (*n* = 220), OV (*n* = 412), and GBM (*n* = 154), (b) Single-end Sequencing for CRC (*n* = 228) and UCEC (*n* = 316). Tumor-lineage-specific pseudogenes were identified based on supervised analysis (ANOVA, corrected *P*-value < 0.05, >1.5 fold change). The color intensity represents the expression level (red, high; blue, low).



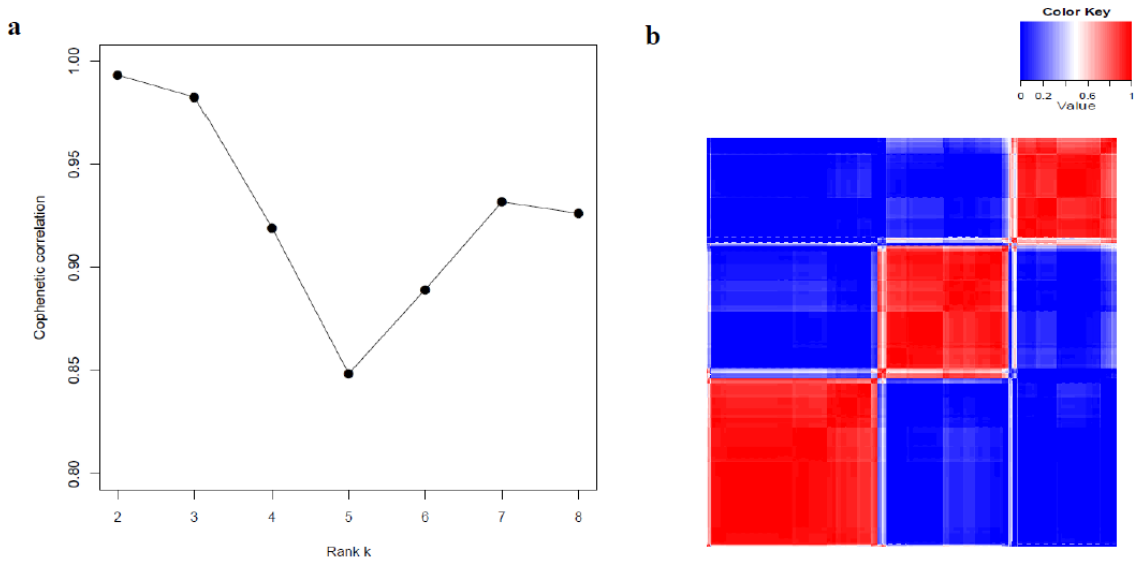

**Supplementary Figure 5. Differentially expressed pseudogenes between normal and tumor samples**

(a) BRCA (pink, *n* = 837) and breast non-tumor (cyan, *n* = 105) samples. (b) KIRC (pink, *n* = 448) and kidney non-tumor (cyan, *n* = 67). (c) LUSC (pink, *n* = 220) and lung non-tumor (cyan, *n* = 17) samples. (d) Overlap of cancer-related differentially expressed pseudogenes in (a)-(c). Cancer-specific pseudogenes were identified based on supervised analysis (ANOVA, corrected *P*-value < 0.05, >1.5 fold change). The color intensity represents the expression level (red, high; blue, low).



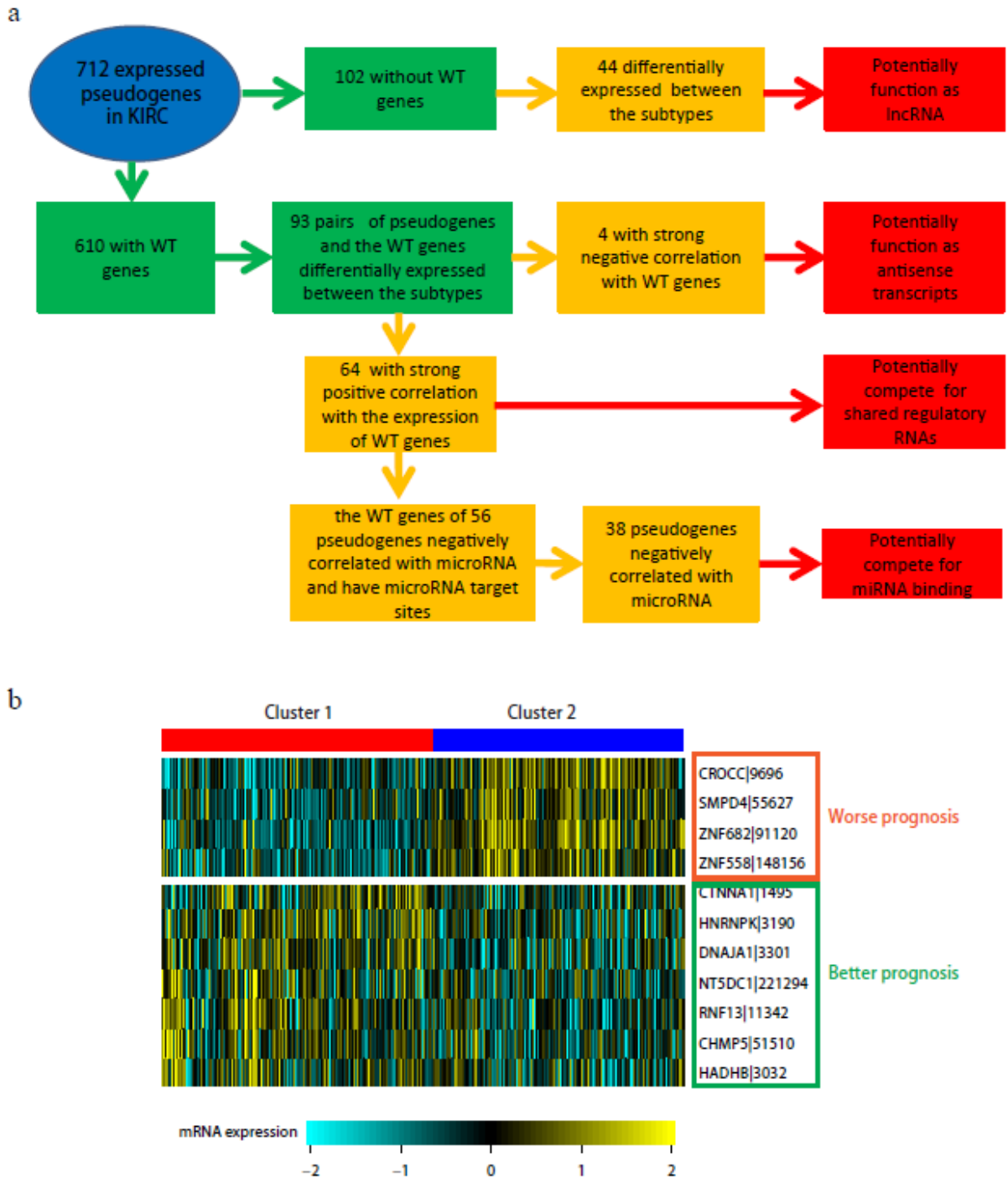

**Supplementary Figure 6. The predictive power of mRNA expression in classifying endometrial subtypes**

(a) The ROC curves of the three classifiers based on the cross-validation within the training set (RF: random forest, SVM: support vector machine, LR: logistic regression.). (b) The ROC curve by applying the best-performing classifier (RF) built from the whole training set to the test set.




**Supplementary Figure 7. Non-negative matrix factorization (NMF) consensus cluster in breast cancer samples**

(a) Distribution of cophenetic correlation based on the number of clusters (K) =2 to 8, that running multiple ranks consecutively can allow for the comparison between differing numbers of clusters using cophenetic correlation. (b) Heatmap for four clusters based on NMF consensus clustering ($n$ = 942).




**Supplementary Figure 8. Potential functional mechanisms of expressed pseudogenes in KIRC**

(a) The analysis summary for inferring the mechanisms through which the expressed pseudogenes may function. (b) Heatmap of up-regulation of better prognosis gene and down-regulation of worse prognosis genes in the pseudogene-expression subtype (subtype 1, with a better-survival, $n$ = 234).