**Description of Additional Supplementary Files**

File Name: Supplementary  Data 1
Description: **FACETS purity**
FACETS purity estimation by sample for ICGC dataset

File Name: Supplementary Data 2
Description: **Differential gene expression analysis for both ICGC & TCGA cohort**
689 genes which are statistically significant in both datasets (glm). P values are adjusted for multiple comparisons

File Name: Supplementary Data 3
Description: **Differential gene expression analysis for both ICGC & TCGA cohort - without copy number log ratio as a covariate**
550 genes which are statistically significant in both datasets (glm). P values are adjusted for multiple comparisons

File Name: Supplementary Data 4
Description: **TCGA Hallmark pathway enrichment score clustering**
Normalised Hallmark pathway enrichment score clusters for the 166 TCGA samples

File Name: Supplementary  Data 5
Description: **CIBERSORTx imputations - statistical comparisons**
Comparison of CIBERSORTx imputations per cell subset between WGD status categories. Kruskal-wallis test p values adjusted using Benjamini-Hochberg method.

File Name: Supplementary Data 6
Description: **CIBERSORTx imputations - raw data**
CIBERSORTx imputations for 78 samples (data with p≥0.05 excluded)

File Name: Supplementary Data 7
Description: **Characteristics of cases for snRNAseq**
Clinical characteristics of cases for snRNAseq

File Name: Supplementary Data 8
Description: **snRNAseq cell annotations**
Single nuclei RNA sequencing final cell clusters, WGD status, aneuploidy prediction, and cell annotations

File Name: Supplementary Data 9
Description: **Proportions of cell types recovered in snRNAseq**
Proportions of cell types recovered in snRNAseq by WGD status

File Name: Supplementary Data 10
Description: **CellChat ligand-receptor pairs**
Statistically significant interactions inferred by CellChat, which internally adjusts for multiple comparisons (bonferroni method). Threshold for statistical significance is p <0.05.