

1 **Description of Additional Supplementary Files**

2

3

4

5

6 Supplementary Data 1: File contains the barcodes of the TCGA tumors used in the study

7

8 Supplementary Data 2: File contains the names of the CCLE cell lines used in the study

9

10 Supplementary Data 3: File contains the top 500 most variable genes by interquartile range for  
11 each tumor type. The GO analysis for these genes are provided in the subsequent sheets.

12

13 Supplementary Data 4: File contains the full pair-wise correlation matrix of CCLE cell lines  
14 compared to TCGA tumors across all tumor types.

15

16 Supplementary Data 5: File contains the ovarian cell line rankings from the Domcke et al.  
17 publication compared to the expression-based rankings derived from this study.

18

19 Supplementary Data 6: File contains the subtype predictions for 9 tumor types. Each sheet  
20 contains cell line subtype predictions, the p-value and FDR for each prediction, and the  
21 expression-based ranking of the cell line compared to the primary tumors in each subtype.

22

23 Supplementary Data 7: File contains a list of the TCGA-110 cell lines, their median correlations  
24 with their respective primary tumors, their predicted subtype if available, and their tumor type.