

Additional file 2

**CTpathway: a CrossTalk-based pathway enrichment analysis
method for cancer research**

Haizhou Liu[†], Mengqin Yuan[†], Ramkrishna Mitra[†], Xu Zhou, Min Long, Wanyue Lei, Shunheng Zhou, Yu-e Huang, Fei Hou, Christine M. Eischen^{*}, Wei Jiang^{*}

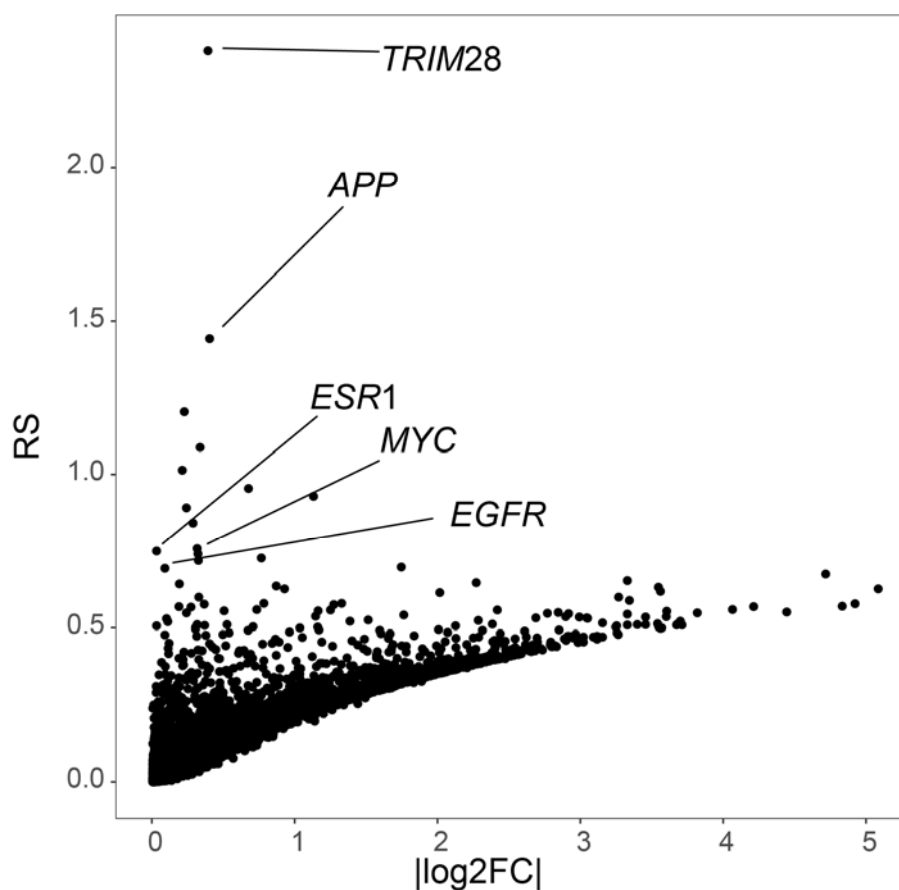


Figure S1. Crosstalk effect consideration detects genes known to be associated with LUAD. Dot plot of genes for GSE116959 LUAD data set. Examples for well-known biomarkers with high *RS* value and low $|\log_2FC|$ value are marked.

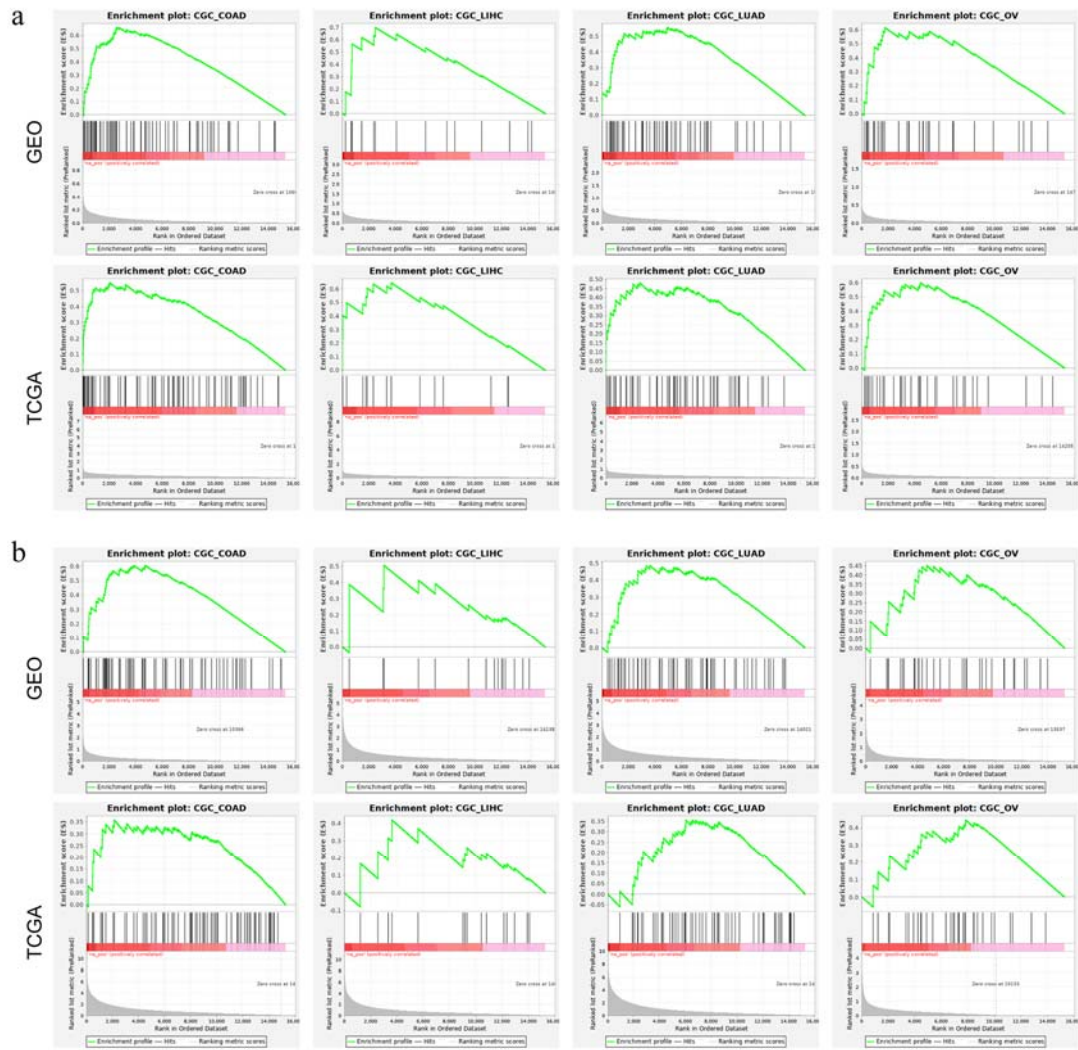


Figure S2. CGC genes are in top of the RS rank list, but not top of the $|\log_2FC|$ rank list. The location of CGC genes enriched in (a) RS rank list and (b) $|\log_2FC|$ rank list. Genes are ranked from high to low according to RS or $|\log_2FC|$.

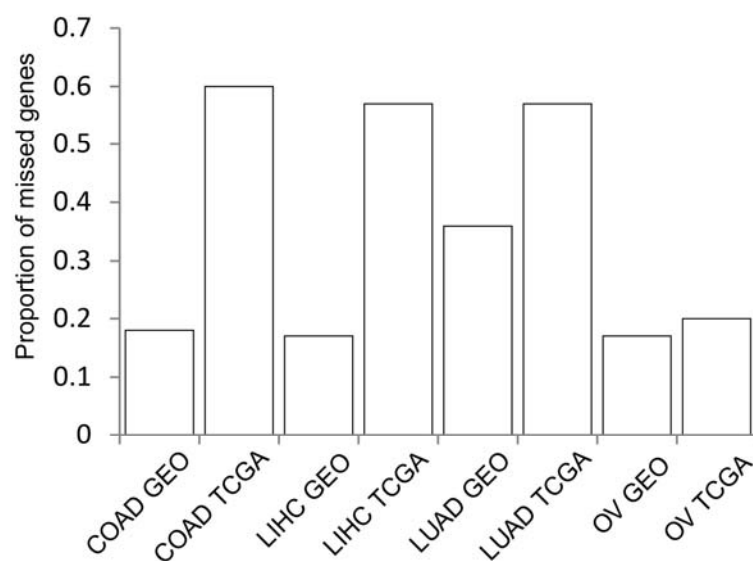


Figure S3. The proportion of risk genes, overlooked by other methods. The proportion of genes with $|\log_2 FC| < 1$ in the top 100 of the *RS* rank list for each dataset.

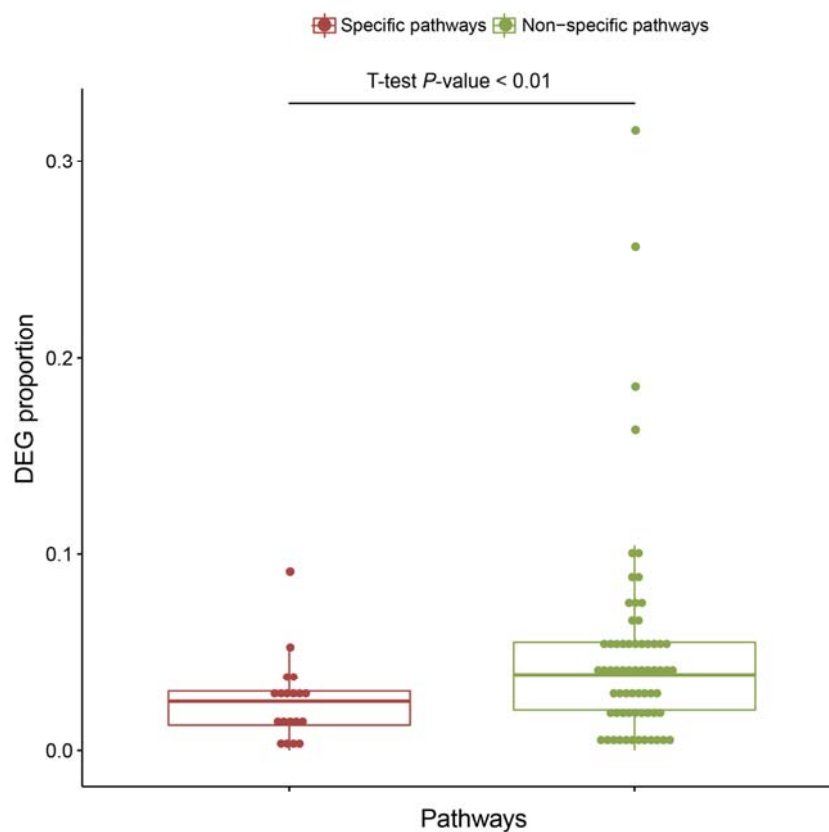


Figure S4. Comparison of DEG proportions between pathways specifically identified by CTpathway and non-specific pathways. The red dots represent specific pathways (identified only by CTpathway); the green dots represent non-specific pathways (identified by both CTpathways and the other methods). The y-axis represents the proportion of DEGs in each pathway.

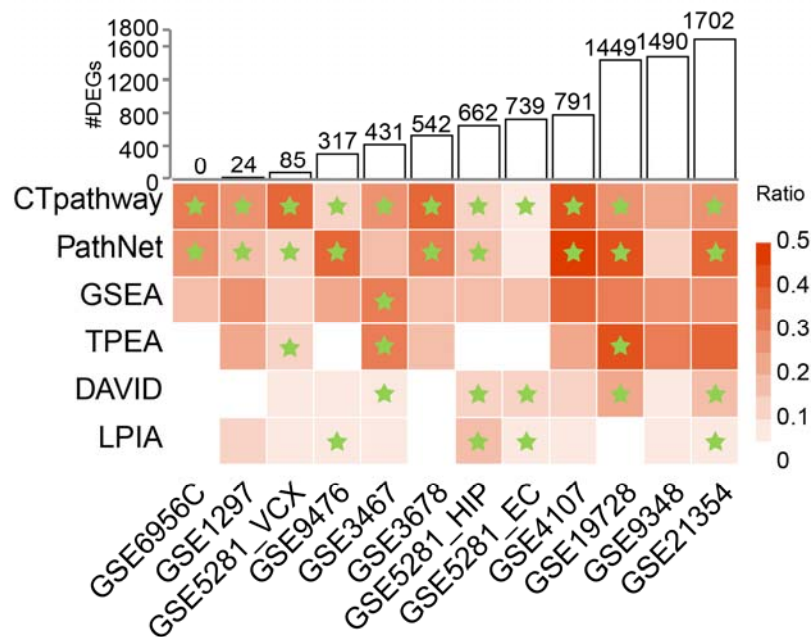


Figure S5. Significant pathways in data sets with a small number of DEGs, identified by CTpathway. The bar graph shows the number of DEGs for 12 representative data sets. The heatmap (below) shows the number of significant pathways identified by different methods at the significant level of P -value < 0.05 divided by the number of all candidate pathways. The target pathways are marked as green stars.

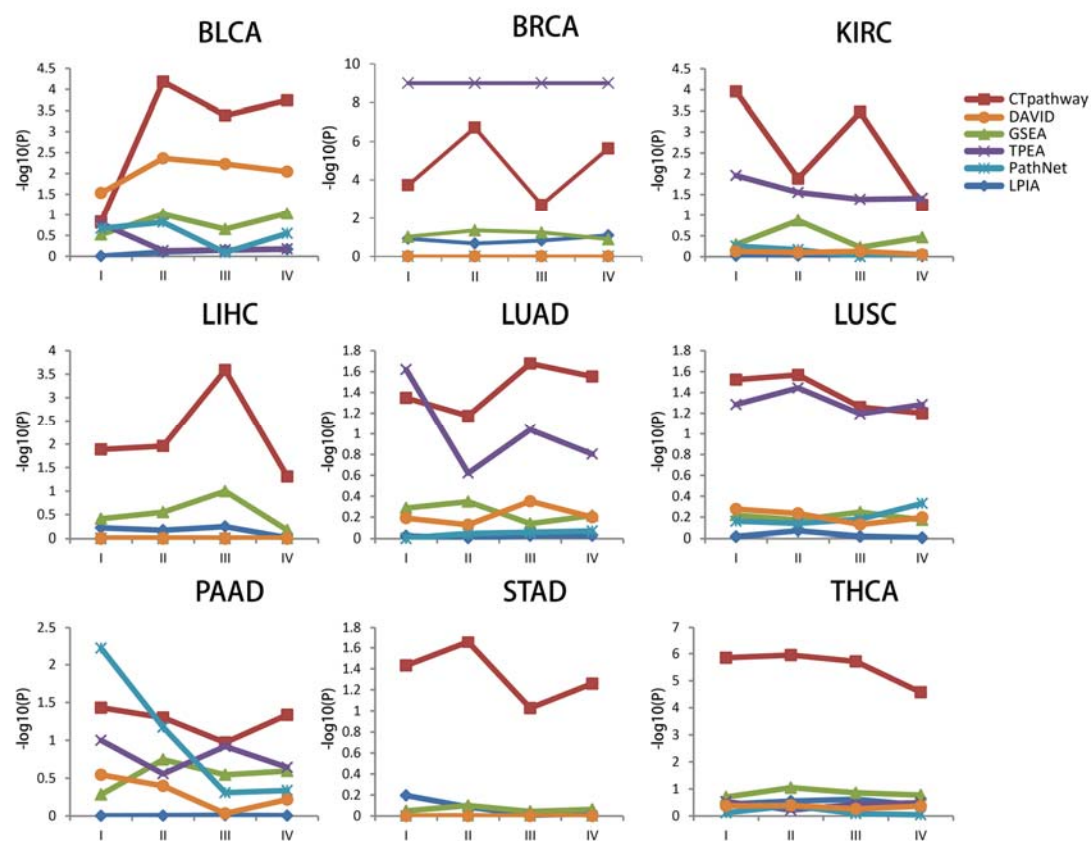


Figure S6. Comparison of enrichment result (P -value) of target pathways for TCGA cancer stages (I, II, III and IV), obtained by different methods.

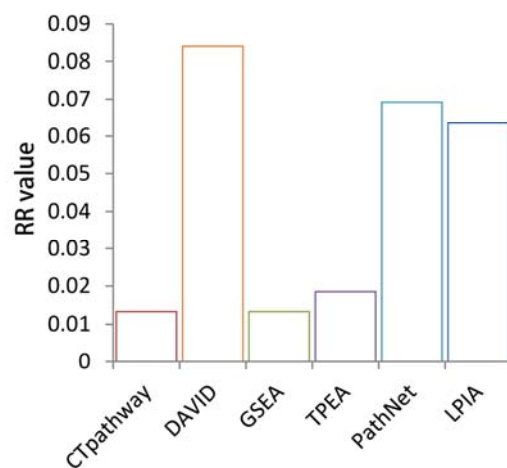


Figure S7. Enrichment result (*RR* value) comparison of target pathways for B cell by different methods.

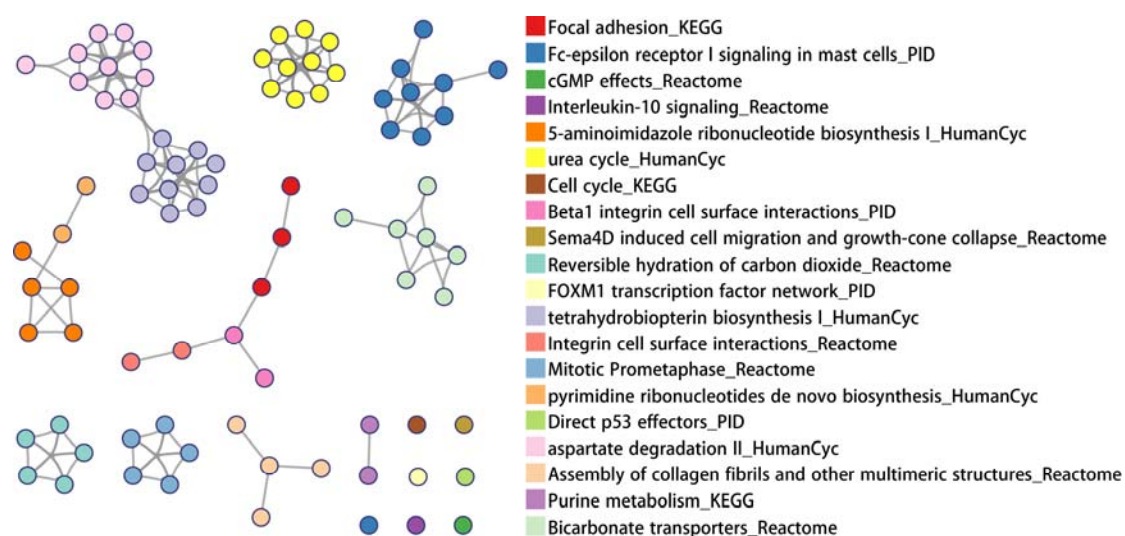


Figure S8. An enrichment map constructed from early-stage COAD data set. The top 20 non-redundant pathways or clusters with low *FDR* determined with CTpathway. For each cluster, the top 10 pathways with low *FDR* are shown in the enrichment map if more than 10 pathways were within one cluster.

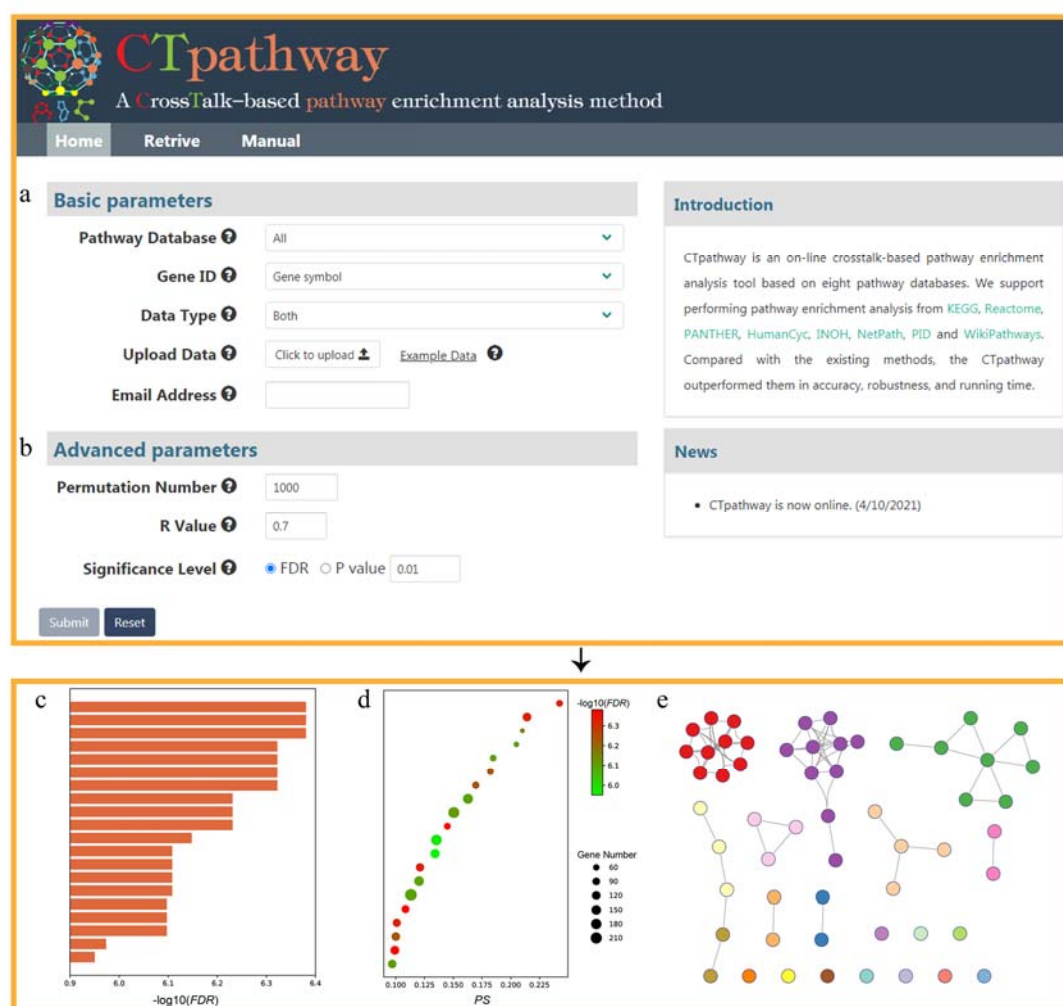


Figure S9. The publicly available CTpathway web tool. (a) Basic parameters input. (b) Advanced parameters input. (c-e) The visualized results are shown in the bar graph (c), bubble plot (d) and enrichment map (e).