Supplementary Material

**Coalesced multimodal learning for the heterogeneity of gastrointestinal cancer based on transcriptome and tissue image data**

Dongmei Ai a,1\*, Yang Dua,1, Hongyu Duanb,1,Lulu Chena, Juan Qi a, Yuduo Wanga, Yishu Wanga, Xiuqin Liua, Li C. Xiab,\*

**\* Correspondence:**

Dongmei Ai

Email Address: [aidongmei@ustb.edu.cn](mailto:aidongmei@ustb.edu.cn) Telephone: 18910789245

Li C. Xia

Email Address: lcxia@scut.edu.cn Telephone: 18602062260

# Supplementary Tables

**Supplementary Table 1.** **Sample sizes of transcriptome data in gastric and colorectal cancer**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample** | **Stomach** | | | **Colon** | | |
| raw data | mRNA | miRNA | | mRNA | | miRNA |
| 414 | 497 | | 524 | | 467 |
| intersection | 376 | | 376 | 445 | 445 | |

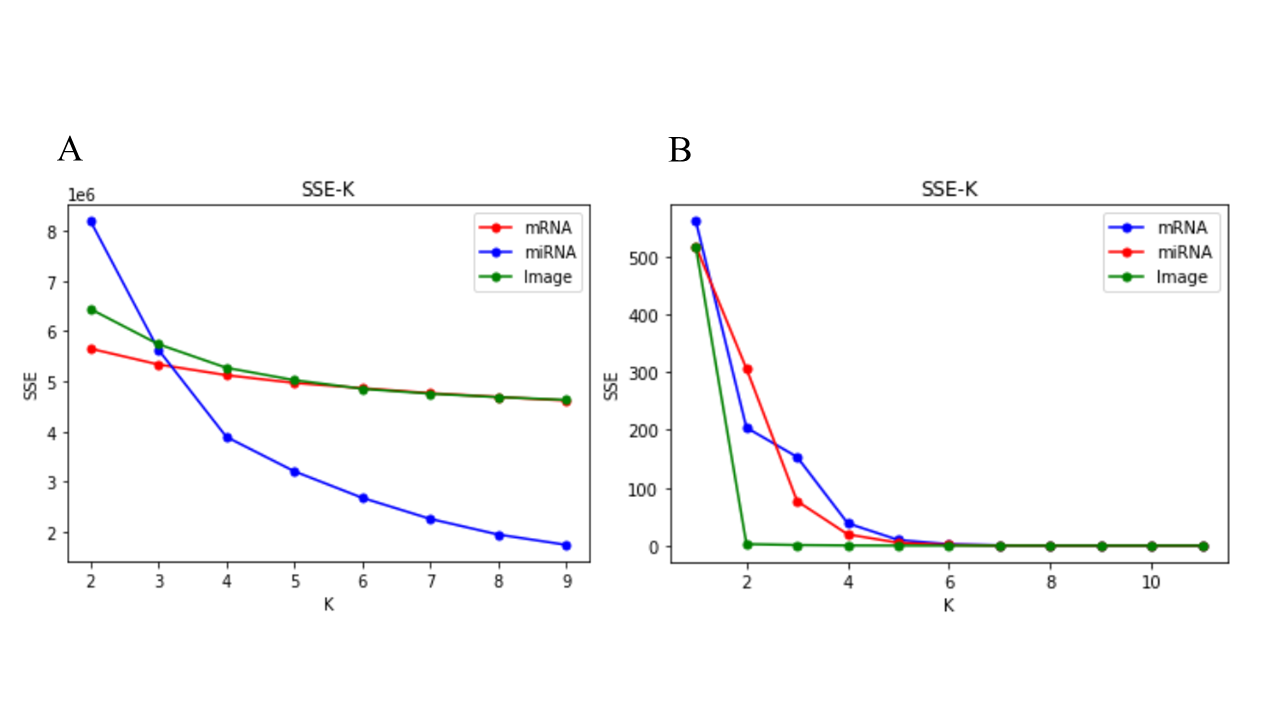
**Supplementary Table 2.** **Features of transcriptome data in gastric and colorectal cancer**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Transcriptome | **Stomach** | | | **Colon** | | |
| raw data | mRNA | miRNA | | mRNA | | miRNA |
| 55812 | 2652 | | 54710 | | 2652 |
| intersection | 54696 | | 2652 | 54696 | 2652 | |

**Supplementary Table 3.** **The number of genes in eight out of 15 modules**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Modules | black | blue | tan | cyan | green | brown | red | grey |
| Number of genes | 229 | 876 | 87 | 70 | 262 | 448 | 232 | 424 |

# Supplementary Figures



Supplementary Fig. 1 A is the SSE (sum of the squared Euclidean distances of each point to its closest centroid)-K curve for the three data views, mRNA, miRNA, and image. The SSE values for mRNA and image were scaled by 0.8, while the SSE value for miRNA was scaled by to fit into the plot. This scaling ensured the convergence of SSE values from all three perspectives into a unified plot, while preserving the cluster count. B is the Curves depicting SSE values and iteration numbers for the similarity matrices associated with the three data views: mRNA, miRNA, and image. These curves encapsulate the evolution of SSE throughout the iteration process.