

Description of Additional Supplementary Files

- Supplementary Data 1:** Datasets used in this study for benchmarking and case study.
- Supplementary Data 2:** Evaluation of the false-positive test in lung cancer.
- Supplementary Data 3:** Benchmarking of early metastatic cell groups — EMT score.
- Supplementary Data 4:** Benchmarking of early metastatic cell groups — CNV score.
- Supplementary Data 5:** Benchmarking of early metastatic cell groups — stemness score.
- Supplementary Data 6:** Proportion of all tools' early metastatic cell clusters for benchmarking.
- Supplementary Data 7:** Benchmarking of early metastatic cell groups — metastasis-related pathway enrichment score.
- Supplementary Data 8:** Metastasis-related pathway enrichment score in negative-metastasis breast cancer.
- Supplementary Data 9:** Overview score across each cancer type.
- Supplementary Data 10:** Early metastasis-related gene signatures across pan-cancer.
- Supplementary Data 11:** EMT score in negative lymph-node data.
- Supplementary Data 12:** Metastasis-related pathway enrichment score in negative lymph nodes (breast cancer).
- Supplementary Data 13:** Gene signatures specific to precursor and occult metastatic cell populations in breast cancer.
- Supplementary Data 14:** Accuracy of predicting high versus low metastatic potential using train/test splits (9:1).
- Supplementary Data 15:** Survival stratified by metastatic potential using attention scores of HSP90AA1 and HSP90AB1.
- Supplementary Data 16:** Accuracy of predicted metastatic time using a leave-one-dataset-out approach.
- Supplementary Data 17:** Survival stratified by metastatic potential using the attention score of ACTB.
- Supplementary Data 18:** TF centrality scores in gene regulatory networks of early-metastasis cell groups across 16 breast cancer datasets.
- Supplementary Data 19:** Transwell results for YY1, JUND and FOXP1 knockouts.
- Supplementary Data 20:** Stemness genes across all cancer types.
- wers' comments (separate file).