Pan-cancer Discriminative Genomic Features Predict Histologically Unknown Tissue Origin of Metastatic Tumors

Biaobin Jiang[1,2,3], Mingrui Jiang[1], Jiguang Wang[1,2,#]

1. Division of Life Science,

2. Department of Chemical and Biological Engineering,

3. Institute of Advanced Study,

The Hong Kong University of Science and Technology, Hong Kong, China

Email: jgwang@ust.hk (J.W.)

Metastasis, a devastating stage of cancer progression, is attributed to around 90% of cancer deaths. Notably, metastatic tumors from different origins preferentially colonize in one distant organ other than uniformly disseminate through circulatory system. This phenomenon, termed organotropism, was discovered over a century ago, but its underlying molecular mechanism is still elusive. In this study, we first aim to identify discriminative genomic features of organotropic metastatic tumors against primary tumors, and then utilize those features to trace the origin of a metastatic tumor with histologically unknown primary site. Concretely, we built a novel computational framework to extract discriminative features from the clinical and genomic information of around 30,000 cancer patients with tumors originating from 30 different primary organs. Comparison of the genomic data between primary and metastatic tumors revealed that metastatic tumors in brain, regardless of the origins, have a significantly higher chromosomal instability than the primary tumors. And the liver metastasis of breast cancers have enriched ESR1 mutations at its ligand-binding domain. Highly frequent alteration of the PI3K/AKT/mTOR pathway in brain metastases of lung cancer leads to their high sensitivities of the targeted therapies. Using those discriminative genomic features, we construct a machine learning model to predict the primary site of metastatic tumors which cannot be determined by histopathological images. As a result, our model achieved an AUPR of 0.44, around 3 fold higher than that of a prior knowledge baseline model. Beside guidance of targeted therapy, this study extends the value of clinical sequencing data in terms of precision diagnosis of metastatic tumors with occult primary.