Computational methods for understanding tumor immune microenvironment from spatial imaging data

Multiplex spatial imaging (MSI) data, such as multiplex immunofluorescence (mIF) or imaging mass cytometry (IMC), allows us to study multiple proteins' activity levels in a tissue section all at once. Unlike traditional methods like immunohistochemistry (IHC), which looks at one protein at a time, MSI technologies provide the opportunity for detailed, multi-parameter spatial descriptions of any area in the context of its immunological landscape. MSI is particularly important in the field of molecular immuno-oncology, where the focus is on understanding the tumor immune microenvironment (TIME). This involves exploring complex interactions between cancer and immune cells and how they influence the tumor development, diagnostics, and therapeutic success. Answers to these questions are crucial for developing effective immunological cancer therapies. Therefore, we formulate three main research goals to unveil TIME and its potential for translation into clinical practices.

Enhancing Precision in MSI Data. MSI data, a key tool in capturing cell identity together with its location, often suffers from errors during generation. These errors can mislead researchers by suggesting false information about protein markers in immune cells or distorting signal strengths. Our first goal is to develop automated methods for data quality control and signal correction. By doing so, we aim to improve the accuracy of MSI data, ensuring that the computational molecular analysis is based on reliable information and free from potential disruptions caused by artifacts.

Unraveling Spatial Complexity in Tissue Organization. MSI data's strength lies in its ability to provide insights into cellular organization within tissues. However, analyzing this data poses computational challenges. To conduct meaningful cohort analyses, we need advanced techniques that reduce data dimensionality while preserving spatial relationships between cells. Therefore, our second aim is to create a suite of graph-based methods. These methods will extract interpretable features, offering a low-dimensional representation of the original MSI data. By expressing clinically significant tumor spatial features using graph theory and image analysis, we aim to introduce new perspectives for TIME classification and understanding.

Bridging Gaps in Multi-Panel MSI Data for Clinical Inference. In our final aim, we plan to develop a statistical model for clinical inference from MSI data. The basic model will integrate data from MSI panels from selected technology for each patient, allowing cohort-wide inference of clinical traits in a Bayesian setting. The final version of the model will incorporate other experimental data modalities, enhancing its efficacy and robustness in predicting clinical outcomes. Importantly, our goal is not only to uncover connections between experimental and observed clinical features but also to infer novel characteristics of cancer patients based on their molecular landscape. This research aligns with our commitment to advancing computational methodologies and gaining deeper insights into the molecular complexities of cancer.

In conclusion, MSI data stands as a crucial yet intricate source of molecular insights, demanding advanced approaches for extraction and processing. Through this project, our aim is to provide effective computational models and methods rooted in graph-driven spatial feature selection and Bayesian probabilistic graphical models for statistical data analysis, that will enable us to answer fundamental questions in the field of cancer immunology. These strides are poised to not only enhance the realms of molecular biology and computer science but, most importantly, foster a deeper comprehension of TIME. This understanding, in the larger context, holds the potential to influence the advancement of more refined diagnostic and therapeutic approaches in the field of oncology and immunotherapy.