

Interpretable multi-omics integration with UMAP embeddings and density-based clustering

Pol Castellano-Escuder¹, Derek K. Zachman^{1,2}, Kevin Han¹,
Matthey D. Hirschey^{1,3,4}

¹Duke University School of Medicine, Duke Molecular Physiology Institute; ²Duke University School of Medicine, Duke Department of Pediatrics/Division of Hematology-Oncology; ³Duke University School of Medicine, Department of Medicine; ⁴Duke University School of Medicine, Department of Pharmacology and Cancer Biology

Integrating high-dimensional multi-omics data is essential for understanding the different layers of biological control. Single-omics methods offer useful insights but often miss the complex relationships between genes, proteins, and metabolites. In this talk, I will present GAUDI (Group Aggregation via UMAP Data Integration), a non-linear, unsupervised method that uses independent UMAP embeddings to analyze multiple data types together. GAUDI reveals relationships across omics layers better than several current methods. It not only clusters samples by their multi-omics profiles but also identifies key features contributing to each cluster, providing clear and interpretable visualizations. I will discuss how GAUDI enables researchers to identify meaningful patterns and potential biomarkers across diverse omics types.

Keywords: Multi-omics, Data integration, UMAP, Clustering, Feature selection.