先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	計算システムズ生物学 (金谷　重彦　(教授))		
学籍番号	2111323	提出日	令和 4年 12月 26日
学生氏名	YANG ZIWEI		
論文題目	An Unsupervised Learning Framework for Identifying Cancer Subtypes		

要旨

Defining and separating cancer subtypes is essential for facilitating personalized therapy modality and prognosis of patients. The definition of subtypes has been constantly recalibrated due to our deepened understanding. During this recalibration, researchers often rely on the clustering of cancer omics data to provide an intuitive visual reference that could reveal the intrinsic characteristics of subtypes. However, while existing studies have shown promising results, they suffer from issues associated with omics data: sample scarcity and high dimensionality. As such, existing methods often impose unrealistic assumptions to extract useful features from the data while avoiding overfitting spurious correlations. In this thesis, we propose to leverage a recent strong generative model, Vector Quantized Variational AutoEncoder (VQ-VAE), to tackle the data issues and extract informative latent features that are crucial to the quality of subsequent clustering by retaining only information relevant to reconstructing the input. VQ-VAE does not impose strict assumptions; hence, its latent features are better representations of the input, capable of yielding superior clustering performance with any mainstream clustering method. We use multiple transcriptomic data (including the expression level of genes and microRNAs) comprising 10 distinct cancers, based on which extensive experiments and medical analysis demonstrate the VQ-VAE clustering results can significantly and robustly improve prognosis over prevalent subtyping systems.