Overlooked Covariates in Metabolite Abundance Levels: Systematically Quantifying the Information Overlap Between Gene Expression and Metabolism Across Multiple Cancer Types

Cong, Thomas (School: Ossining High School)

Dysregulated metabolism is a hallmark of cancer progression yet exactly how cancer cells vary their metabolic program during tumorigenesis remains unknown. Moreover, metabolism is highly dynamic, which makes it challenging to study experimentally. Previous computational approaches have utilized gene expression to infer metabolic profiles but failed to consider effects of external covariates. In this study, we sought to systematically quantify the relationship between gene expression, metabolome, and covariates. We developed and applied a novel computational framework that combines unimodal covariate analysis with multimodal distance correlation to integrate and analyze data from the Cancer Atlas of Metabolic Profiles and the Cancer Cell Line Encyclopedia. Four confounding covariates were identified in cancer cell line data, and these make substantial contributions to the perceived information overlap between transcriptome and metabolome which masks the “true” relationship between metabolite levels and metabolic gene expression. Unimodal analyses associated 20.2% of metabolic variance with tissue of origin and 4.7% with extracellular environment, while 8.2% of variance associated with growth rate. In real tumor data, tumor purity serves as the major covariate, similar to cell line data where immunological cell lines have distinct metabolic profiles. This underscores precautions that must be taken into consideration when comparing metabolic cell line data to those from real tumors. Results of this study provide a standardized procedure for quantifying the complex relationship between transcriptome and metabolome. Furthermore, these methods have the potential to significantly advance our understanding of metabolic drug resistance and immune response.