

Integrative image and RNA-seq data analysis**Momiaio Xiong**

University of Southern California, USA

Abstract

Emerging integrative analysis of increasingly larger and complex genomic and anatomical imaging data which has not been well developed, provides invaluable information for the holistic discovery of the genomic structure of disease and has the potential to open a new avenue for discovering novel genetic influence on the imaging variation and disease which cannot be identified if they are analysed separately. Both imaging and genomics generate a huge amount of data that present critical bottleneck in the interactive analysis. A key issue to the success of imaging and genomic data analysis is how to reduce dimensions of both imaging and genomic data. Most previous investigated methods for imaging information extraction and RNA-seq data reduction do not explore imaging spatial information and often ignore gene expression variation at genomic positional level. To overcome these limitations, we extend one dimensional principle component analysis to two dimensional principle component analyses (2DFPCA) and use 2DFPCA scores to represent image data. Unlike the widely used linear regression for modelling association of microarray gene expressions with phenotypes which quantifies the expression level of a gene/transcript by a single number, we develop a multiple functional linear model (MFLM) in which functional principal scores of images are taken as multiple quantitative traits and RNA-seq profile across a gene is taken as a function predictor for assessing the association of gene expression with imaging signals which can take gene splicing and expression variation at genomic positional level into account. The developed method has been applied to image and RNA-seq data of ovarian cancer and KIRC studies. We Identified 24 and 84 genes whose expressions were associated with imaging variations in ovarian cancer and KIRC studies, respectively. Our results showed that many significantly associated genes with images were not differentially expressed, revealed their morphological and metabolic functions. The results also demonstrated that the peaks of the estimated regression coefficient function in the MFLM often allowed discovering splicing sites and multiple isoform of gene expressions.

Received: June 6, 2022; **Accepted:** June 13, 2022; **Published:** June 20, 2022**Biography**

Momiaio Xiong is a Professor at Division of Biostatistics and Human Genetic Center, School of Public health, the University of Texas Health Science Center at Houston. He obtained his PhD degree from the Department of Statistics

at the University of Georgia in 1993 and finished his postdoctoral training in computational biology from the University of Southern California in 1995. His research is in the area of statistical genetics, bioinformatics, machine learning and image analysis.