



The 7th Conference on Bioinformatics, 3-5 January 2018

Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran



Stratification of breast cancer subtypes using somatic mutation and none negative matrix factorization

Mohsen Falahat^{a*}, Javad Hassanpour^b

^a Faculty of new science and technology, University of Tehran, Tehran, Iran

^b Department of Information and Communication Technology, Malek Ashtar University Of Technology, Tehran, Iran

* mohsenfalahat@ut.ac.ir

Abstract: Breast cancer is the second leading cause of cancer death of women worldwide. It is a heterogenous disease with multiple subtypes which have different causes and clinical outcomes. Stratification of tumors is one of the major tasks in cancer genomics for a better understanding of the tumor heterogeneity and better targeted treatments. Somatic mutation data provides a rich new source of data for identifying these subtypes. In this work we use somatic mutation data and none negative matrix factorization to uncovering breast cancer subtypes. Also we use Jaccard distance to generate a new profile. The major benefit of using Jaccard distance is that it reduces the genetic heterogeneity and sparsity. We verified the accuracy of the proposed method by showing meaningful associations between the resulting subtypes and clinical features. Results showed that our method can identify clinically meaningful tumor subtypes. it has high predictive power for some clinical features includes histological type, ER status, PR status and tumor stage.

Keywords: somatic mutation; tumor stratification; NMF