[BIOINFO 2023]

Discovering environment-based disease similarity network using NHANES datasets

Younghoon Kim[1]

[1]*Department of Supercomputing Applications,*
Korea Institute of Science and Technology Information (KISTI)

Disease incidences are caused by both of genetic and environmental factors. However, many studies have only focused on genetic factors causing diseases so far. In order to understand disease incidence mechanism well, we need to investigate environmental causes and the relationship between genetic and environmental factors. Here we studied environmental factors for many diseases by analyzing NHANES datasets (National Health and Nutrition Examination Survey). NHANES datasets are the national survey data to investigate people in US for public health management purposes. These data contain laboratory examination results including many environmental factors in human blood and urine. By using these datasets, first we identified the environmental factors causing each disease. We performed logistic regression for 40 diseases and 300 environmental factors. After identifying the causing environmental factors, we performed clustering analysis to see the environment-based human disease similarity network. In the results, we could find that mouth is the hub node of this similarity network and the similarity network is similar to physiological pathways.