[BIOINFO 2024]

Analysis of the influence of genetic and lifestyle factors on gout using machine learning

Do-Hyeon Kwak[1], Hyunjung Kim[1], Sun Shim Choi[1*], and Ki-Won Moon[2*]

[1] *College of Biomedical Science, Institute of Bioscience & Biotechnology,*
Kangwon National University
[2] *Division of Rheumatology, Department of Internal Medicine,*
Kangwon National University School of Medicine
**Corresponding author:* [1]schoi@kangwon.ac.kr, [2]kiwonmoon@kangwon.ac.kr

Gout is a form of inflammatory arthritis with a complex etiology that involves both genetic and lifestyle factors. Genome-wide association studies (GWAS) have identified numerous genetic loci significantly associated with gout. However, most gout-related GWAS have been conducted in European populations. Our study aimed to determine whether gout-associated single-nucleotide polymorphisms (SNPs) identified in several European GWAS were applicable to a Korean cohort. By calculating polygenic risk scores (PRS) for the Korean cohort based on SNPs from these European GWAS, we observed that the gout group had higher PRS scores than the non-gout group in the Korean cohort as well. Subsequently, we developed a machine learning (ML) model to distinguish between gout and non-gout individuals using lifestyle habits and PRS values. The study included 43,988 non-gout control samples and 617 gout case samples from the KoGES cohort. To address the significant imbalance between control and case samples, ML was performed using 100 random down-samplings of the control samples. The model achieved an AUC of 0.66, and we observed a slight improvement in the model performance when PRS values were included. Overall, PRS calculation using SNPs from European GWAS is applicable to the Korean cohort and may be useful for predicting of gout occurrence.