Universität Münster

Institut für Medizinische Informatik IMI

medizinische fakultät
Universität Münster

Titel:

Detecting sample mix-ups in omics and clinical data

Abstract:

Patients are often characterized by multiple data types, including diverse omics and clinical data. To ensure quality of downstream analyses, it is important that samples are correctly paired. However, practice shows that accidental sample mix-ups regularly occur.
Various tools can automatically detect possible mix-ups between omics data based on genetic relatedness, but novel methods are needed that also support clinical data.
As omics and clinical data can not be directly compared, the first challenge lies in estimating clinical traits from genetic data. Traits like sex are easily imputed, but can only detect mix-ups between patients of opposite sex and are not sufficient to correct those mix-ups. Other traits like bloodtype or height can be estimated on the basis of SNPs, either directly or using machine learning.
The second challenge is to develope a scoring system that represents the likeliness of reported and estimated clinical data belonging to the same patient. Low scoring pairs indicate possible mix-ups and alternative pairings can be suggested that result in higher scores.