Published online: 2020-08-21

Content Summaries of Best Papers Selected for the IMIA Yearbook Section on Bioinformatics and Translational Informatics

Béal J, Montagud A, Traynard P, Barillot E, Calzone L

Personalization of logical models with multi-omics data allows clinical stratification of patients

Front Physiol 24 Jan 2019;9:1965

This paper presents a new methodology - called PROFILE - to adapt generic logical models of cancer pathways to a particular biological sample (*i.e.,* patient's tumor). The authors use as a proof of concept a published model of cancer signaling pathways of breast cancer. The approach integrates mutation data, CNA (copy number alterations), and expression data to personalize the logical model to a patient's profile. The simulation of the resulting models (using the MaBoSS program) shows a good correlation with patient's subgrouping on NPI (Nottingham Prognostic Index) and survival time. This paper illustrates the potential of using logical modeling (and concepts of systems biology) for precision medicine as it can eventually facilitate the choice of patient-specific drug treatment thanks to a self-explanatory model.

Chen ML, Doddi A, Royer J, Freschi L, Schito M, Ezewudo M, Kohane IS, Beam A, Farhat M

Beyond multidrug resistance: Leveraging rare variants with machine and statistical learning models in Mycobacterium tuberculosis resistance prediction

EBioMedicine 2019 May;43:356–69

Multidrug-resistant tuberculosis (MDR-TB) is still a public health challenge because of the lengthy current culture-based antimicrobial susceptibility testing due to in vitro growth of *mycobacterium tuberculosis* (MTB). As an alternative to recent molecular tests for MDR-TB criticized for low sensitivity and the small number of tested drugs, the authors propose a whole-genome sequencing approach to capture both common and rare mutations responsible for drug resistance. They use a neural architecture combining a wide and deep neural network (WDNN) compared to simpler classifiers such as logistic regression and random forests. They leverage whole-genome sequencing of the pathogen, as well as rare variants and known drug resistant variants to predict the resistance to 10 anti-tuberculosis drugs. They achieved AUCs over 93% for first-line drugs, and 89% for second-line drugs.

Kim K, Baik H, Jang CS, Roh JK, Eskin E, Han B

Genomic GPS: using genetic distance from individuals to public data for genomic analysis without disclosing personal genomes

Genome Biol 2019 Dec;20(1):175

Genomic global positioning system (GPS) applies the multilateration technique commonly used in the GPS to genomic data. This framework allows to calculate genetic distances from considered samples to reference samples (public data), and share this information with others. This sharing enables certain types of genomic analysis, such as identifying sample overlaps, close relatives, mapping to geographical origin without disclosing personal genomes. This innovative approach allows achieving a good balance between open genomic data sharing and privacy protection.

Marttinen M, Paananen J, Neme A, Vikram M, Takalo M, Natune T, Paldanius KMA, Mäkinen P, Bremang M, Kurki MI, Rauramaa T, Leinonen V, Soininen H, Haapasalo A, Pike I, Hiltunen M

A multiomic approach to characterize the temporal sequence in Alzheimer's disease-related pathology

Neurobiol Dis 2019;124:45468

This paper proposes a multi-omics and temporal sequence-based approach to provide a better understanding of the sequence of events leading to Alzheimer's Disease (AD). The authors coupled transcriptomic and phosphoproteomic data to determine the temporal sequence of changes in microRNA, protein, and phosphopeptide expression levels from human temporal cortical samples, with varying stages of the AD. This approach highlighted a significant sequence of key functions occurring at the considered stages of the disease, namely: *(i)* fluctuation in synaptic and mitochondrial functions as the earliest pathological events in brain samples with AD-related pathology, and *(ii)* the increased expression of inflammation and extracellular matrix-associated gene products. The authors made use of decision trees and random forests for identifying potential biomarkers predicting the disease degree.