

Temos pavadinimas: Multi-omikos duomenų integravimas imunologijoje: mokslinei bendruomenei skirtos duomenų bazės kuravimas ir mašininio mokymosi modelių kūrimas įgyto ir įgimto imuniteto tyrimams

Title of the topic: Multi-omics data integration for immunology: building a curated community database and machine learning models for adaptive and innate immunity research

Aprašymas

Nors pastaraisiais metais multi-omikos metodų pažanga paskatino atradimų bangą imunologijoje, skirtingų biologinių lygmenų duomenų integravimas išlieka esminiu iššūkiu. Dabartinės duomenų bazės paprastai apima tik vieną omikos sluoksnį - metabolomiką, transkriptomiką arba genomiką, todėl visapusiška daugiamatė analizė tampa neįmanoma. Dėl to vertingi tyrimų duomenys lieka išsibarstę ir sunkiai prieinami. Šiuo projektu siekiama išspręsti minėtas problemas - mokslinei bendruomenei bus sukurta nauja, specialiai imunologijai pritaikyta duomenų bazė, apjungianti konsorciumo surinktus metabolomikos, lipidomikos, transkriptomikos ir genomikos duomenis. Duomenų rinkiniams bus taikoma griežta kokybės kontrolė, jie bus susieti su metaduomenimis ir pakoreguoti pašalinant eksperimentinių grupių efektą (angl. batch effect). Tai užtikrins didesnę statistinę galią bei patikimus įgimto ir įgyto imuniteto tyrimus. Remiantis šia duomenų baze, bus kuriami specifiniams audiniams pritaikyti statistiniai ir mašininio mokymosi metodai, skirti kokybinei bei kiekybinei multi-omikos duomenų analizei. Statistiniai modeliai padės identifikuoti ląstelių populiacijas ir lipidų grupes, kurias paveikė liga ar taikytas gydymas, o prižiūrimo mokymosi (angl. supervised learning) algoritmai leis išskirti specifinius biožymenų rinkinius. Visi šie analizės procesai bus integruoti į bendrą platformą - tai bus lengvai prieinamas, lanksčiai konfigūruojamas analizės įrankis, kuriam naudoti nereikės specializuotų bioinformatikos žinių. Galutinis šio projekto rezultatas - unikali tarpdisciplininė platforma, paspartinsianti naujus atradimus imunologijoje.

Abstract:

Recent advances in multi-omics have driven significant discoveries in immunology, yet integrating heterogeneous datasets remains a major bottleneck. Existing databases typically focus on a single omics layer, such as metabolomics, transcriptomics, or genomics, limiting comprehensive cross-dimensional analysis. For this reason valuable datasets remain scattered and inaccessible. This project addresses these gaps by building a novel, immunology-specific community database that consolidates metabolomic, lipidomic, transcriptomic, and genomic data generated across the consortium. Datasets will be curated for quality, harmonised with metadata, and corrected for batch effects to enable robust adaptive and innate immunity research with greater statistical power. Building on this resource, the project will develop tissue-specific computational, statistical, and machine learning methods for analysing qualitative and quantitative multi-

omics data. Statistical models will identify cellular clusters and lipid groups affected by disease or drug treatments, while ensembles of supervised learning models will derive discriminative biomarker panels representing cellular and disease-specific fingerprints. These validated workflows will be implemented within a unified platform as an accessible, parameter-driven analysis tool requiring no bioinformatics expertise, automatically generating transparent, reproducible reports. The outcome is a unique, interdisciplinary resource that accelerates immunological discovery, empowering researchers in academia and industry worldwide.

Mokslinis vadovas / supervisor: Dr. Edita Karosienė, Head of Science at VUGENE

Kontaktai / Contacts:

El. p. / Email: edita.karosiene@vugene.com