Project title: Developing explainable data integration of omics and imaging data in cardiovascular diseases

Principal investigator: Dr. Rachel Cavill

Promotor: Prof. Dr. E.A.L. (Erik) Biessen,

Collaborators: Dr. J.M.H. (Joël) Karel

Proposal (250 words):

Background:
When multiple sets of data are collected in parallel about each patient (for instance gene expression, histopathological images and metabolite information), then the integration of these datasets is complex and challenging, even more so as the number of parameters is high and the number of subjects low. This project will develop new methods and algorithms to implement data integration strategies for such data and apply them with our collaborators in the cardiovascular field.

Study objective:
In this project the PhD will develop integrative data analysis techniques to
(1) Develop explainable machine learning which combines the advantages of multivariate data analysis (PCA/PLS variants) with enrichment methods (eg GSEA), to produce interpretable results when integrating multiple omics datasets from matched samples.
(2) Develop new algorithms for the integration of data extracted from spatial spectral and histological images.
(3) Apply the algorithms developed in (1 and 2) to find reproducible, reliable associations between datasets both public and locally generated from the field of cardiology.
The student will work as part of an international, multidisciplinary team of bioinformaticians, computer scientists and biologists, at the departments of Advanced Computing Sciences and Pathology, having strong interactions with leading international research groups in systems medicine, data analysis, mass spectroscopy, and metabolomics.

Requirements candidate: The project requires knowledge in mathematics, bioinformatics or computer science with a focus on data science and/or machine learning. Interest in medical applications. Good English and communication skills.

Keywords: data mining; multivariate statistics; explainable machine learning; data fusion; data integration; bioinformatics.

Top 5 selected publications:
1. H. Jin, P. Goossens, P. Juhasz, W. Eijgelaar, J. Karel, E. Smirnov, C. Sikkink, B. Mees, O. Waring, K. van Kuijk, G. Fazzi, M. Gijbels, M. Kutmon, C. Evelo, U. Hedin, M. Daemen, J. Sluimer, L. Matic, E. Biessen, Integrative multiomics analysis of human atherosclerosis reveals a serum response factor-driven network associated with intraplaque hemorrhage, Clin. and transl. med., 2021, 11(6):e458.
2. R. Cavill, D. Jennen, J. Kleinjans, JJ Briede. Transcriptomic and metabolomic data integration. Briefings in Bioinformatics, 2016, 17 (5) 891-901.
3. A. Kamburov, R. Cavill, T. Ebbels, R. Herwig, H. Keun, Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA, Bioinformatics, 2015, 27(20):2917-2918.
4. B. Legein, L. Timmerman, E. Biessen, E. Lutgens, Inflammation and immune system interactions in atherosclerosis, Cellular and molecular life sciences, 2013, 70:3847-3869.
5. Cavill, R., Kamburov, A., Ellis, J. K. J. K., Athersuch, T. J. T. J., Blagrove, M. S. C. M. S. C., Herwig, R., … Keun, H. C. Consensus-Phenotype Integration of Transcriptomic and Metabolomic Data Implies a Role for Metabolism in the Chemosensitivity of Tumour Cells. PLoS Computational Biology, 2011, 7(3), 12.