

## Machine Learning Based Classification of Meningiomas to Clinically Relevant Molecular Subtypes using Methylation Profiling

### Authors:

*Tomi Viitanen<sup>1,2</sup>, Juha Kesseli<sup>1,2</sup>, Elaheh Moradi<sup>1,2</sup>, Kirsi Rautajoki<sup>1,2</sup>*

<sup>1</sup> Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

<sup>2</sup> TAYS Cancer Centre, Tampere University Hospital, Tampere, Finland

### Keywords:

cancer, bioinformatics and computational biology, machine learning, molecular diagnostics, epigenetics

### Abstract

Meningiomas are the most common primary central nervous system tumors, with approximately 80% considered benign. The diagnosis into WHO grades is solely based on histopathology, which remains the standard, but is prone to inaccuracies due to interobserver variability and inclusion of criteria lacking validation. Moreover, WHO grades alone offer limited prognostic value, as substantial heterogeneity exists within tumors of the same grade. In response, Nassiri et al. (2021) has identified clinically relevant molecular subtypes for meningiomas. Notably, from a cohort of 1039 histopathologically benign meningiomas 26% were classified into high-risk molecular subtypes. We aim to develop a machine learning classifier for these molecular subtypes utilizing a select set of features from DNA methylation microarrays for targeted sequencing or PCR-based diagnostic approaches.

We used Illumina HumanMethylation EPIC array data from 824 meningioma samples, 400 of which had molecular subgroup labels, to train a machine learning-based classifier. Two methylation feature subsets were identified from the data: 1) features showing intraclass consistency and variation between classes and 2) features explaining grouping variation independent of sample subtype labels. Semi-supervised learning will be used to iteratively label unclassified samples to increase the amount of training data. The final classifiers are trained with different sized groups of the most informative methylation features from the subsets. Performances of the classifiers will be assessed on an independent validation set.

Preliminary results show that a support vector machine using 2314 selected features yields 100% classification accuracy on samples with moderately high confidence score, and an overall accuracy of 93%.