

Radiotherapy biomarkers discovery using machine learning approaches

Björn Andersson^a, Peidi Liu^a, Britta Langen^b, Eva Forsell^b, Marcela Davila^a

^aBioinformatics Core Facility, The Sahlgrenska Academy at the University of Gothenburg, Göteborg, Sweden

^bDept. of Radiation Physics, Inst. of Clinical Sciences, Sahlgrenska Cancer Center, The Sahlgrenska Academy at the University of Gothenburg, Göteborg, Sweden

Aim

The aim of the project is to develop an unbiased machine learning tool/pipeline using different omics data (i.e. transcriptomic and proteomic) to determine a panel of biomarkers that can reflect the dose-dependency for radiation treatments.

Introduction

Both radiation research and cancer radiotherapy heavily depend on the accuracy of the relation among the absorbed dose, radiation type and modality, and the biological effects in healthy or malignant tissues after irradiation. Incorrect estimation of the dose-response causes severe problems of under-treatment or damage to healthy tissues, leading to increased risk of cancer recurrence and reduced quality of life. Therefore, molecular biomarkers that correctly indicate the dose-response of a tissue are needed for reducing the risks.

Results

- Both differential expression analysis of the omics data and the GA/KNN algorithm, identified potential biomarkers.
- Regression with identified biomarkers showed that the predicted radiation doses are well correlated to the true doses (Figure 1).

Conclusion

We found significantly regulated transcripts and shared transcripts between tissues and/or conditions. GA/KNN is a powerful iterative machine learning algorithm that could select a small panel of features (in this case regulated genes) that are potential biomarkers for absorbed dose. This biomarkers can then be evaluated in lab experiments to confirm their functions.

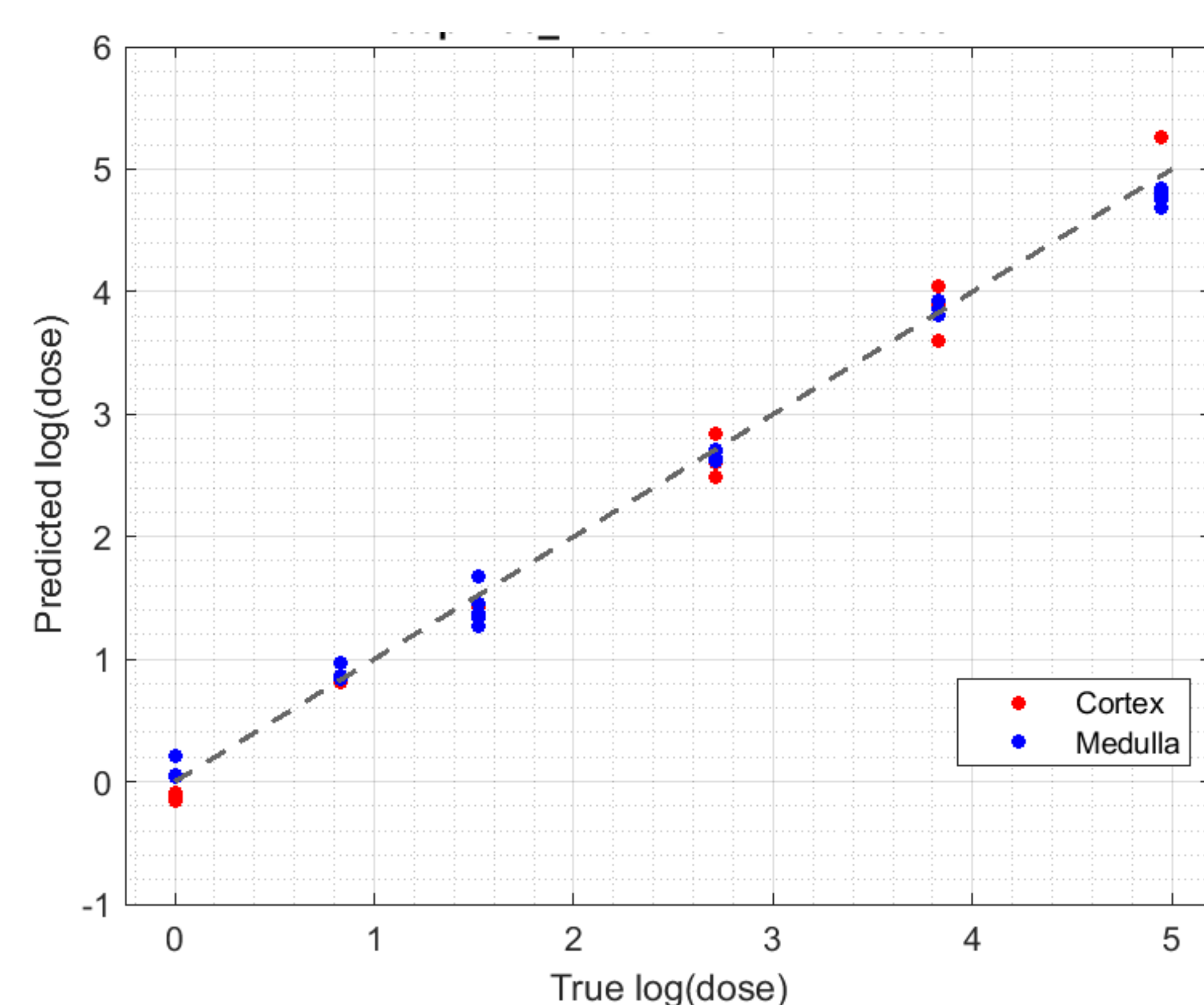


Figure 1. Dose regression with biomarker discovered using GA/KNN algorithm.

Methods

Transcriptomic data (microarray) from animal experiments (mouse) that were treated (i.v. injected) with different radionuclides (iodine-131, astatine-211, ¹⁷⁷Lu-octreotate and ¹⁷⁷Lu-chloride). Absorbed dose over time was calculated according to the MIRD formalism (ref 1) for the following tissues: thyroid, kidneys, liver, lungs, spleen and blood.

Using Genetic Algorithm and K-nearest Neighborhood (ref 2) machine learning pipeline (Figure 2) to select top ranked genes/proteins that could classify radiated samples from untreated controls based on tissue types, absorbed dose and time after treatment.

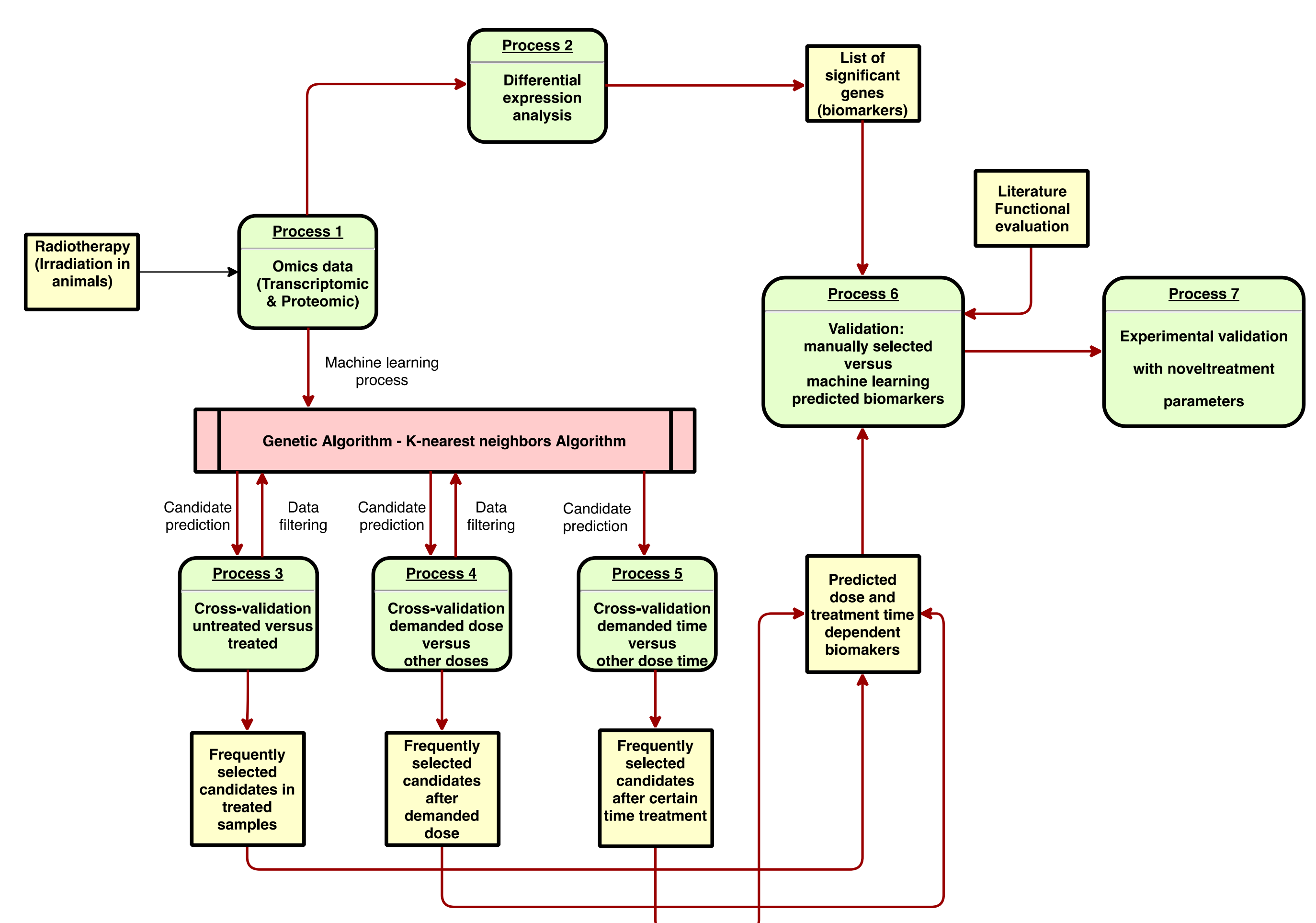


Figure 2. Analysis pipeline for radiotherapy biomarker discovery using GA/KNN algorithm.



References

- Ref 1. Bolch WE, Eckerman KF, Sgouros G, Thomas SR. MIRD pamphlet No. 21: a generalized schema for radiopharmaceutical dosimetry--standardization of nomenclature. J Nucl Med. 2009;50:477–84.
 Ref 2 .Li, L., Darden, T. A., Weinberg, C. R., Levine, A. J. and Pedersen, L. G., 2001a. Gene Assessment and Sample Classification for Gene Expression Data Using a Genetic Algorithm / k-Nearest Neighbor Method. Comb. Chem. High Throughput Screen 4, 727–739.