JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

Connecting Raman Biomarkers with Breast Cancer Cell Immunohistochemical Types by Semi-supervised Machine Learning Methods

Authors: Xinchen Deng[1], Kirsty Milligan[1], Ramie Ali-Adeeb[1], Jeffrey L. Andrews[2], Phillip Shreeves[2], Julian J. Lum[3,4], Alexandre Brolo[5], Andrew Jirasek[1]
Address: [1]Department of Physics, University of British Columbia – Okanagan, [2]Department of Statistics, University of British Columbia – Okanagan, [3]University of Victoria, Department of Biochemistry and Microbiology [4]Trev and Joyce Deeley Research Centre [5]University of Victoria, Department of Chemistry

INTRODUCTION

In the current research, we demonstrate that Raman spectroscopy in combination with a semi-supervised non-negative matrix factorization and random forest combined(ssNMF-RF) model can be used to predict and track cellular immunohistochemical type for a panel of cell lines exposed to varying doses of ionizing radiation. Furthermore, we also show that individual biochemical bases can be identified and attributed to decreasing orders of relevance in predicting cellular immunohistochemical status. The dataset includes one normal breast cell line MCF10A and four breast cancer cells (MCF7, BT474, MDA-MB231, and SKBR3), and the cells were exposed to varying doses of single fractions of radiation (0, 10, 30, 50Gy). Raman spectra of the reference biochemical base chemicals were collected for subsequent input into the ssNMF-RF model. Using the novel ssNMF method, we obtained scores for each Raman biomarker within our interest from different Raman reference libraries. A random forest classifier was then fitted to the biochemical scores for performing immunohistochemical type classifications (HER2, PR, ER, Ki67, and cancer vs non-cancer) and selecting important biomarkers for each classification. Overall, the ssNMF-RF model have yielded classification results with high accuracy (> 0.9), high sensitivity (> 0.9), and high specificity (> 0.9). The feature selection from random forest was able to rank how much each biochemical contributed in each immunohistochemical information classification test.

METHOD

Radiation delivery and cell culture

- One normal breast cell line MCF10A and four breast cancer cells (MCF7, BT474, MDA-MB231, and SKBR3),
- Radiation delivery with a Varian 21EX linear accelerator at varying doses of single fractions of radiation (0, 10, 30, 50Gy).

Raman spectroscopy

Raman spectra acquisition of cells: inVia Raman microscope (Renishaw Inc., Gloucestershire, UK), a 100X dry objective (NA = 0.9), a 1200 lines / mm diffraction grating, a 10 second acquisition time per cell, and a 785 nm laser (Renishaw).

Figure 1. Raman spectroscopy machine from the Jirasek lab

ssNMF-RF model scheme

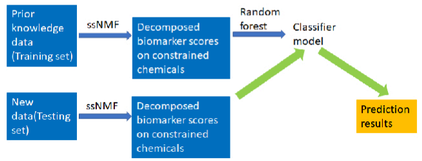

Figure 2. ssNMF and random forest combined classifier model training and data workflow

RESULTS

Summary of results: Predict immunohistochemical profiles of the breast cell lines

Cell lines	HER2	PR	ER	Ki67	Subtype	Cancer or not
MCF10A				*low*	*Basal*	*no*
MCF7	–	+	+	*low*	*LuminalA*	*yes*
MDA MB231				*high*	*Basal like*	*yes*
BT474	+	+	+	*high*	*LuminalB*	*yes*
SKBR3	+	–	–	*high*	*HER2*	*yes*

Table 1. Immunohistochemical profiles of the breast cell lines

- The aim of the project is to connect Raman biomarkers with immunohistochemical profiles of breast cell lines. The Raman breast cellular samples were first decomposed in ssNMF with constrained chemicals, resulting chemical scores for each corresponding chemical.
- Due to the complexity of the chemical scores, a random forest was then applied on the score to classify the scores with regard to the immunohistochemical types.
- Random forest can also feature selection to select important chemicals for each classification.

Random forest for classification

Table 2. Immunohistochemical types and cancer vs non-cancer classification results

Types	Accuracy		Sensitivity		Specificity	
	mean	S.D	mean	S.D	mean	S.D
HER2	0.968	0.011	0.976	0.010	0.976	0.010
PR and ER	0.968	0.009	0.980	0.013	0.980	0.013
Ki67	0.993	0.006	0.999	0.002	0.999	0.002
Cancer or not	0.994	0.004	0.971	0.020	0.971	0.020

Select important chemicals: feature selection with random forest

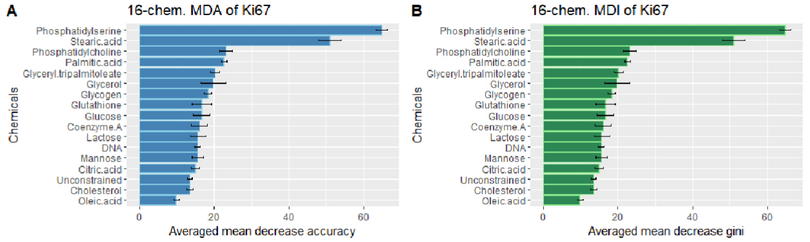

Figure 4. Variable importance of Ki67 test with Random forest feature selection

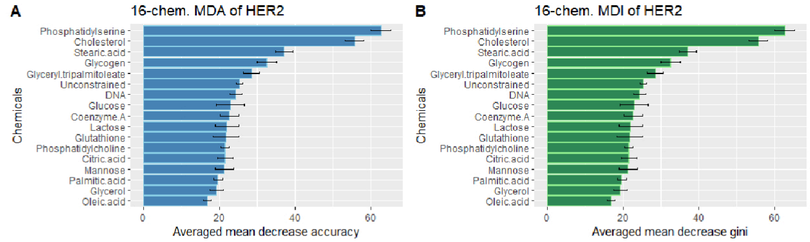

Figure 5. Variable importance of HER2 test with Random forest feature selection

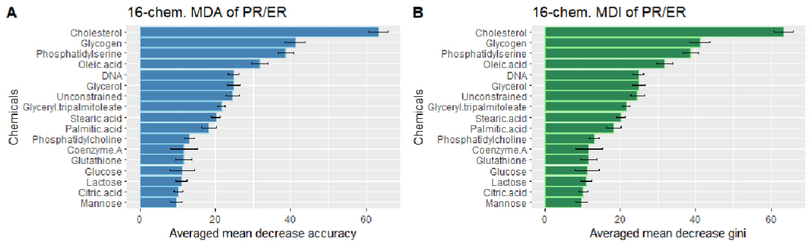

Figure 6. Variable importance of PR/ER test with Random forest feature selection

CONCLUSIONS

- The ssNMF-RF classifier model had preliminary success to connect Raman biomarkers to immunohistochemical profiles of different breast cancer cell lines.
- The ssNMF and random forest approach had classification results for immunohistochemical profiles with high accuracy(>0.9), high sensitivity(>0.9), and high specificity(>0.9).

REFERENCES

Meksiarun P, Aoki PHB, Nest SJV, Sobral-Filho RG, Lum JJ, Brolo AG, et al. Breast cancer subtype specific biochemical responses to radiation. Analyst. 2018;143(16):3850–3858. doi:10.1039/C8AN00345A.

Matthews Q, Jirasek A, Lum J, Brolo A. Biochemical signatures of in vitro radiation response in human lung, breast and prostate tumour cells observed with Raman spectroscopy. Physics in medicine and biology. 2011;56:6839–55. doi:10.1088/0031-9155/56/21/006.

Matthews Q, Isabelle M, Harder SJ, Smazynski J, Beckham W, Brolo AG, et al. Radiation-Induced Glycogen Accumulation Detected by Single Cell Raman Spectroscopy Is Associated with Radioresistance that Can Be Reversed by Metformin. PLOS ONE. 2015;10(8):e0135356. doi:10.1371/journal.pone.0135356.

ACKNOWLEDGEMENTS

This work was supported by grant funding from Natural Sciences and Engineering Research Council and Canadian Institutes for Health Research.
The authors declare that there is no conflict of interest.

CONTACT INFORMATION

alison.deng@ubc.ca