Classifying Breast Cancer Cell Lines of Different Metastasis Potentials using Visible Resonance Raman Spectroscopy and Machine Learning

Binlin Wu¹, Kenneth Jimenez², Lin Zhang³, Susie Boydston-White⁴, Eric Wang¹,⁴, Cheng-hui Liu², and Robert R. Alfano²

¹Physics Department and CSCU Center for Nanotechnology, Southern Connecticut State University, 501 Crescent Street, New Haven, CT 06515, USA.
²Institute for Ultrafast Spectroscopy and Lasers, Department of Physics of the City College of the City University of New York, 160 Convent Avenue, New York, NY 10031, USA.
³Science Department, Borough of Manhattan Community College of the City University of New York, 199 Chambers Street, New York, NY 10007-1097, USA.
⁴Amity Regional High School, 25 Newton Rd, Woodbridge, CT 06525

wu1@southernct.edu

ABSTRACT

About 1 in 8 U.S. women (about 12%) will develop invasive breast cancer over the course of her lifetime. In 2020, an estimated 276,480 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S., along with 48,530 new cases of non-invasive (in situ) breast cancer. About 42,170 women in the U.S. are expected to die in 2020 from breast cancer. Besides skin cancer, breast cancer is the most commonly diagnosed cancer among American women. In 2020, it's estimated that about 30% of newly diagnosed cancers in women will be breast cancers.

Optical biopsy: fluorescence spectroscopy, Raman spectroscopy, multiphoton imaging, SERS, CARIS, ...

- Rapid
- Non-invasive or minimally invasive
- Diagnostic information for early detection (biological, morphological)
- Avoid subjectivity

VISIBLE RESONANCE RAMAN SPECTROSCOPY

- Signal enhancement due to resonance or pre-resonance.
- Signal enhancement due to use of shorter wavelength: Raman scattering cross section is inversely proportional to the fourth power of the excitation wavelength.
- Resonance large biomolecules: flavins, lactate, NADH, NAD+, collagens, elastin, carotenoids, tryptophan, heme proteins, mitochondrial cytochromes.

SAMPLES AND EXPERIMENTS

Instrument and experimental parameters
- WiTec alpha300 Raman microscope system, equipped with a Nikon 20X objective.
- Excitation: 532 nm, 3.5mW on the sample, spot size 1 μm
- Spectral range: 200 cm⁻¹ to 4000 cm⁻¹

Samples
- In total, 68 spectra acquired: 32 normal, 18 non-aggressive, 18 aggressive
  - MCF-10A: Healthy or benign (control) | fibrocystic disease
  - MCF-7: Non-aggressive | Ductal carcinoma
  - ER+, PR+, HER-
  - Luminal breast cancer cell line
  - MDA-MB-231: Very aggressive | Ductal carcinoma
  - ER+, PR+, HER- (Triple-negative breast cancer)
  - Cannot be treated by traditional hormonal methods

RESULTS

- Non-negative Matrix Factorization (NMF)
  \[ X = WH \]
  - Principal Component Analysis (PCA)
  - PCA is a unsupervised machine learning method to maximizing variance explained in X.
  - PCs are the components that characterize the data most strongly
  - W: columns are PC spectra
  - H: rows are PC scores used for classification

CONCLUSIONS

- Using VRR and PCA-SVM or NMF-SVM may be able to detect the aggressiveness of breast cancer cells with ≥80% accuracy.
- This approach may be used for early detection of the metastatic breast cancers, especially triple negative breast cancer (TNBC).

FUTURE WORK

- Increase sample size
- Look into the key Raman peaks that help classify the aggressiveness
- Further improve detection accuracy (e.g. use deep learning)
- Evaluate nuclear and DNA information