INTRODUCTION

- Current methods to detect and manage cervical pre-cancers often miss the disease and generate false positives.
  - This can lead to a delay in correctly diagnosing and over-treating cervical dysplasia.
  - Recent clinical trials with two year follow up have shown that colposcopy (i.e. current triage standard of care) can be inaccurate in determining the need for biopsies and locating appropriate biopsy sites.

- Guided Therapeutics has developed a spectroscopic device to reduce the number of false positives while retaining the high true positive predictive rate.

- Traditional supervised and unsupervised methods of analyzing fluorescence and reflectance spectra data have difficulty recovering physiological information.
  - Supervised unmixing methods require knowledge about the reflectance/fluorescence spectral patterns of the constituent materials.
  - Reflectance and fluorescence spectral data are non-negative by nature, and spectral sources don’t always appear to be statistically independent (e.g. collagen and elastin fluorescence).
  - However, principal component analysis assumes that the underlying sources are uncorrelated and Gaussian, and the resulting source/concentrations can be negative.
  - Also independent component analysis assumes that the sources are statistically independent, and the sources/concentrations can be negative.

  Therefore, non-negative matrix factorization (NMF) is chosen to analyze the spectral data, since it only assumes that the source and mixture data are non-negative.

METHODS

1. Clinical colposcopic expert and biopsy results are used to create a point-by-point diagnosis.
   - The diagnostic categories include: normal (e.g. normal squamous, normal columnar), CIN 1 (Grade 1), CIN 2 (Grade 2), CIN 3 (Grade 3), or CIN 2+ (either CIN 2 or CIN 3).

2. Measurement points and subjects are excluded if they were taken over non-cervical tissue, had excessive blood or mucus, had a specular reflection artifact (CCD saturation), or had an unknown/uncertain disease classification.

3. Spectral measurements are calibrated to account for any instrument and subject differences.
   - Calibration sources are used to calibrate the system’s spectra response and perform wavelength calibration.
   - Subject differences are calibrated by dividing the spectra with the mean intensity over all wavelengths.

4. A NMF approach is blindly decompone fluorescence and reflectance spectroscopy into a set of constituent source curves and concentrations.
   - NMF approach with the product of two other non-negative matrices: X=A*S where X is the measured spectral data matrix, S is an unknown spectral source matrix, and A is an unknown mixing matrix.
   - Matrices S and A are chosen to minimize the root-mean-squared residual (RMSR) between X and A*S.
   - NMF algorithm is an iterative approach and does not reach a unique solution.

5. Lasso regression machine learning modeling is used to identify the combination of different source concentrations that predict the amount of cervical dysplasia.
   - The Lasso is a shrinkage and selection method for linear regression that minimizes the sum of squared errors with a bound on the sum of the absolute values of the coefficients.
   - The quantitative prediction labels used in the Lasso model are 1 for normal points, 2 for CIN 1, 3 for CIN 2, and 4 for CIN 3.
   - The model is randomly divided into 60% training and 40% testing data to test for over fitting.
   - A receiver operating characteristic curve (ROC) analysis is done to determine the dysplasia tissue classification performance.

6. Two dimensional (2-D) disease maps are created to spatially locate and quantify cervical dysplasia tissue using the Lasso regression results obtained from the NMFNMF concentrations.
   - Each subject’s 2-D false color disease map is created from the Lasso regression results and the corresponding measurement location.
   - Finally, the 2-D disease maps are compared with cervical colposcopes images and biopsy results.

RESULTS

- Figure 3. Comparison between derived and actual source reflectance and fluorescence spectra.
  - (a) shows the reflectance spectrum signal that best predicts the diagnosis by itself.
  - (b) shows the fluorescence source signal that predicts cervical dysplasia by itself.
  - The quantitative prediction labels are 1 for normal points, 2 for CIN 1, 3 for CIN 2, and 4 for CIN 3.

NMFNMF sources appear to characterize physiological cervical tissue spectra.

- The behavior of the reflectance spectra is opposite of the absorption spectrum, and you can see the characteristic HBO2 absorption peaks at 660nm as measured in the reflectance/fluorescence source signal.

- The quantitative prediction labels used in the Lasso model are 1 for normal points, 2 for CIN 1, 3 for CIN 2, and 4 for CIN 3.

REFERENCES

- Current methods to detect and manage cervical pre-cancers often miss the disease and generate false positives.
- This can lead to a delay in correctly diagnosing and over-treating cervical dysplasia.
- Recent clinical trials with two year follow up have shown that colposcopy (i.e. current triage standard of care) can be inaccurate in determining the need for biopsies and locating appropriate biopsy sites.
- Guided Therapeutics has developed a spectroscopic device to reduce the number of false positives while retaining the high true positive predictive rate.

- Traditional supervised and unsupervised methods of analyzing fluorescence and reflectance spectra data have difficulty recovering physiological information.
- Supervised unmixing methods require knowledge about the reflectance/fluorescence spectral patterns of the constituent materials.
- Reflectance and fluorescence spectral data are non-negative by nature, and spectral sources don’t always appear to be statistically independent (e.g. collagen and elastin fluorescence).
- However, principal component analysis assumes that the underlying sources are uncorrelated and Gaussian, and the resulting source/concentrations can be negative.
- Also independent component analysis assumes that the sources are statistically independent, and the sources/concentrations can be negative.

Therefore, non-negative matrix factorization (NMF) is chosen to analyze the spectral data, since it only assumes that the source and mixture data are non-negative.

1. Clinical colposcopic expert and biopsy results are used to create a point-by-point diagnosis.
   - The diagnostic categories include: normal (e.g. normal squamous, normal columnar), CIN 1 (Grade 1), CIN 2 (Grade 2), CIN 3 (Grade 3), or CIN 2+ (either CIN 2 or CIN 3).

2. Measurement points and subjects are excluded if they were taken over non-cervical tissue, had excessive blood or mucus, had a specular reflection artifact (CCD saturation), or had an unknown/uncertain disease classification.

3. Spectral measurements are calibrated to account for any instrument and subject differences.
   - Calibration sources are used to calibrate the system’s spectra response and perform wavelength calibration.
   - Subject differences are calibrated by dividing the spectra with the mean intensity over all wavelengths.

4. A NMF approach is blindly decompone fluorescence and reflectance spectroscopy into a set of constituent source curves and concentrations.
   - NMF approach with the product of two other non-negative matrices: X=A*S where X is the measured spectral data matrix, S is an unknown spectral source matrix, and A is an unknown mixing matrix.
   - Matrices S and A are chosen to minimize the root-mean-squared residual (RMSR) between X and A*S.
   - NMF algorithm is an iterative approach and does not reach a unique solution.

5. Lasso regression machine learning modeling is used to identify the combination of different source concentrations that predict the amount of cervical dysplasia.
   - The Lasso is a shrinkage and selection method for linear regression that minimizes the sum of squared errors with a bound on the sum of the absolute values of the coefficients.
   - The quantitative prediction labels used in the Lasso model are 1 for normal points, 2 for CIN 1, 3 for CIN 2, and 4 for CIN 3.
   - The model is randomly divided into 60% training and 40% testing data to test for over fitting.
   - A receiver operating characteristic curve (ROC) analysis is done to determine the dysplasia tissue classification performance.

6. Two dimensional (2-D) disease maps are created to spatially locate and quantify cervical dysplasia tissue using the Lasso regression results obtained from the NMFNMF concentrations.
   - Each subject’s 2-D false color disease map is created from the Lasso regression results and the corresponding measurement location.
   - Finally, the 2-D disease maps are compared with cervical colposcopes images and biopsy results.

CONCLUSIONS

- Calculated sources from the NMF approach appear to be represent physiological sources (e.g. hemoglobin and porphyrin) and their concentration changes are in agreement with cervical dysplasia changes.

- A machine learning algorithm (i.e. Lasso linear regression) improved the dysplasia prediction performance by combining source concentrations from both the reflectance and fluorescence spectra data.

- 2-D false color spectroscopy disease maps demonstrate the ability to quantify and spatially locate dysplasia cervical tissue.

- These results offer the potential to reduce the number of false positive cases while maintaining a high enough detection rate necessary for primary screening, and may assist in identifying biopsy locations.

ACKNOWLEDGEMENTS

The project described was supported in part by a grant award from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

© 2011 Guided Therapeutics, Inc.