Prediction of Cancer-Associated Skeletal Muscle Wasting Using Targeted Profiling of Urinary Metabolites

Roman Eisner*, Jianguo Xia, David Hau, Thomas Eastman, Cynthia Stretch, Sambasivarao Damaraju, Russell Greiner, David S. Wishart, and Vickie Baracos

*Contact: eisner@cs.ualberta.ca

Introduction

•Cancer-associated skeletal muscle atrophy (cancer cachexia)

- Involuntary weight gains or losses are significant perturbations of precise metabolic, neuronal, and hormonal controls
- Associated with poor functional status, treatment toxicity and shorter life expectancy
- •Muscle wasting may be an early or occult phenomenon that is difficult to detect against the background of overall body weight
- Muscle loss may occur independently of changes in fat mass

• Improved approaches to detecting the onset and evolution of muscle wasting would help manage wasting syndromes and facilitate early intervention

• Gold standards for measuring body fat and muscle over time:

- Dual energy X-ray absorptiometry (DXA)
- Computed Tomography (CT)
- Magnetic Resonance Imaging (MRI).
- These methods are expensive, their analysis may be time-consuming and labor -intensive, and they may expose the patients to radiation

• Recent developments in NMR-based metabolomics permit detection and quantification of dozens of metabolites from urine (metabolic profile)

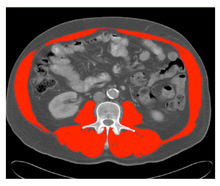
• We use machine learning approaches to build a classifier that can predict muscle loss for novel patients, based on his/her metabolic profile

Data Set

• Study was reviewed and approved by the Alberta Cancer Board Research Ethics Board

• **Patients:**

- 73% had lung (n=66) and 27% colorectal cancer (n=25)
- Donated a spot urine sample
- Body composition assessed by review of several CT images

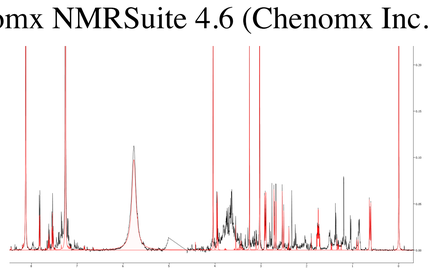


- Total skeletal muscle tissue cross-sectional area (cm^2) at the 3rd lumbar vertebra using Slice-O-Matic software V4.3 (Tomovision, Montreal)
- Muscle area at the CT image preceding the urine sample collection was used as a reference (baseline) to compute the percentage of muscle lost or gained by the next imaging time point

•**Urine samples:**

- One-dimensional NMR spectra of urine samples were acquired
- First increment of the standard NOESY pulse sequence on a four-channel Varian (Varian Inc., Palo Alto, CA) Inova-600 MHz NMR spectrometer with a triax -gradient 5-mm HCN probe

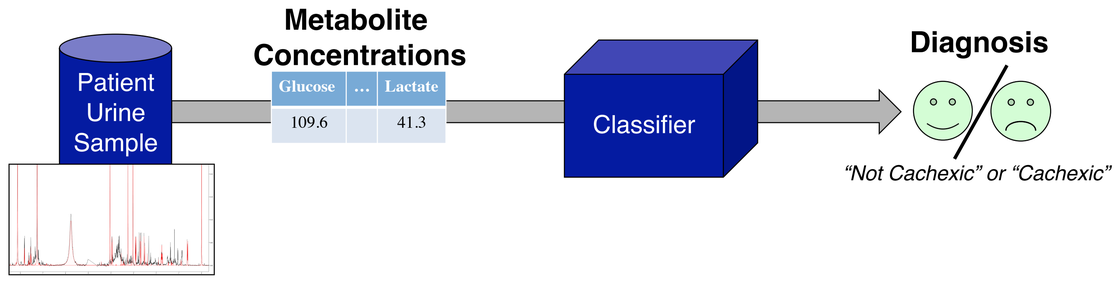
• We use the targeting profiling approach, acquiring the concentrations of 63 metabolites, using Chenomx NMRSuite 4.6 (Chenomx Inc. Edmonton, Canada)



Prediction of Cancer Cachexia

• **Goal:**

- Given a patient's urine sample, predict whether the patient has cachexia



• **Sample Analysis:**

- Metabolite concentrations were log-transformed to make distributions more Normal
- Common approach: just compute *correlation* between outcome (here, cachexia status of patients) with with each individual observed variable (metabolite concentration in urine samples)
- Instead we build a *diagnostic tool* to predict whether patients are cachexic based on metabolic profile (from urine samples)

• **Machine Learning:**

- 1.) *Train* classifier from historical (labeled) data
- 2.) Use classifier to *predict* muscle loss of novel patient
- Evaluated:
  - Novel algorithm, Pathway-Informed Analysis (PIA)
  - PLS-DA (commonly used in metabolomics)
  - Other well known ML/Statistics approaches

• **Pathway-Informed Analysis (PIA)**

- Bayesian classifier
  - Computes P(anabolic | metabolic profile) and P(catabolic | metabolic profile)
    - Returns larger of the two (i.e. the most likely diagnosis)
  - •Issue: How to efficiently model the relationships among the metabolites?

• Use known metabolic pathways to model metabolite relationships:

- Kyoto Encyclopedia of Genes and Genomes (KEGG) provides a database of metabolic pathways in humans
- Include only metabolites appearing in metabolic profile *and* KEGG
- Use these pathways to create the structure of the Guassian Markov Random Field (GMRF):
  - Nodes represent metabolites
  - Edges represent common reactions between metabolites

• PIA performs better than other commonly used approaches

- Permutation testing shows result is significantly better than random

Classifier	5-fold Cross-validation Accuracy
Pathway Informed Analysis	79.3 %
Full dependence model	72.2 %
Support vector machine (SVM)	72.2 %
Naïve Bayes model	71.1 %
PLS – DA	68.1 %
Tree-augmented naïve Bayes	62.6 %
Decision tree	59.7 %
Random permutation	49.9 %