PROOF Centre: Who are we? What Problems are we Tackling?

- Centre for the **Prevention Of Organ Failures** is a non-profit organization established in 2008
- Federally and industry funded to create a world-class national centre of excellence
- Canadian annual expenditures exceed $70 billions on end-stage heart, lung and kidney diseases
- Growing burden due to aging populations
Our Approach: Biomarker Solutions

Distinct biological indicators (cellular, biochemical or molecular) of a process, event or condition that can be measured reliably in tissues, cells or fluids.

Sets of genes and/or proteins in the blood identified using whole genome technologies.
Impacting the Life Cycle of Organ Failures

Management of therapies and risk factors

Rapid decliners

Slow decliners

Biomarker Opportunities

Early detection

End-stage markers

State of Health (%)

Duration (years)
Our Program in Chronic Obstructive Pulmonary Diseases (COPD)

- Over 1.4M Canadians and 14M Americans suffer from COPD
- Exacerbations (“lung attacks”) can be lethal and require extensive hospitalization
- Currently difficult to predict who will progress from no exacerbation to frequent exacerbations
- We are developing a prognostic blood test to help identify “rapid decliners”
Current Management of COPD

High risk
Low risk

Clinical Judgment

"severe"
Over-treated

"mild"
Under-treated

Clinical Judgment

"severe"
Over-treated

"mild"
Under-treated

Clinical Judgment

"severe"
Over-treated

"mild"
Under-treated

Clinical Judgment

"severe"
Over-treated

"mild"
Under-treated
How will a prognostic test change care?

Clinical Judgment

Prognostic Blood Test

High Risk

Low Risk

Appropriate Treatment
Impact of a COPD Exacerbation
Blood Test

Future Costs

Future Deaths

Base case
Smoking cessation
New drugs to enhance lung function
New tests to prevent exacerbation
Working with health care providers, industries and governments to bring these blood tests to patients within 3 years

Hope to obtain FDA approval in 5 years

Cheaper and yet better medicine

What do all these have to do with Outliers and Outlier Explanations??
A. Detecting Potential Labeling Errors

- Biomedical data can be very noisy:
  - Laboratory environment could change
  - Diagnostic decisions are not completely objective
  - Different “gold-standards” are used for grading

- Essential to check for label (e.g., severity of condition) consistency
A. Our Approach [MBN06]

- Propose a leave-one-out perturbed classification matrix:
  - Flip every training sample and compare the resulting classifier with the classifier trained on the original training set
  - A training sample A is a suspect of mislabeling if flipping A’s label significantly increases cross validation accuracy (using SVM)
- Effectiveness shown on 3 real microarray data sets with ground truths
- Identified a few suspicious cases in PROOF’s projects, e.g., timing of blood draws, leading to a journal publication
B. MDQC: Our Approach to Microarray Quality Control [FZN+07]

- A QC report inspects a microarray in isolation
- Goal: to identify outlying arrays that are not evident from inspection of individual arrays
  - High-throughput whole genome technologies process arrays in batches of 96
- collapses all the values in QC reports into measures to assess the quality of the array, the sample, and the RNA
- measures the distances of each array to an “average” array in the study, adjusting for covariances
B. MDQC Advantages

- Performs a *multidimensional* analysis and not requiring absolute thresholds (which are often arbitrary)
- Easy to implement and visualize, and computationally inexpensive (as compared with Affy PLM)
- Can suggest potential sources of problems and possible batch effects
Summary

- Personalized medicine relies critically on very clean data; outlier detection plays a valuable role.
- Whole genome technologies rapidly advancing; but platforms not yet stable; important application domain for outlier detection.
- Finding “intelligent” information associated with outliers help the user to understand the outliers, track down potential problems, and intervene as early as possible.