

Q5 Probabilistic Classification Algorithm

Draft Training Plan

Oregon Health & Science University, Adopters

DOCUMENT CHANGE HISTORY

Version Number	Date	Description
1.0	March 2006	Initial Document

TABLE OF CONTENTS

DOCUMENT CHANGE HISTORY	2
1. INTRODUCTION.....	4
2. APPROACH.....	5
2.1 SELF PACED TUTORIAL MODULES:.....	5
2.2 HANDS-ON TRAINING	5
3. AUDIENCE	6
4. DELIVERY.....	7
4.1 HANDS-ON TRAINING	7
4.2 SELF-PACED LEARNING.....	7
4.3 USER GUIDE	7

1. INTRODUCTION

The purpose of this document is to define the training approach to be carried out by the OHSU Adopter team for the Q5 software project. Q5 is a probabilistic classification algorithm for classification of proteomic mass spectra of cancerous and non cancerous serum samples. Currently Q5 works for SELDI and preprocessed MALDI data. The goal of the training is to familiarize biostatisticians with the format of proteomics data (if necessary) and the Q5 program itself.

2. APPROACH

The training program includes two self-paced learning modules and a hands on session intended to familiarize users with the command-line interface to Q5. As many biostatisticians tend to be self-directed learners, the hands-on session is optional and may be replaced with the second tutorial module.

2.1 SELF PACED TUTORIAL MODULES:

This first part of the training consists of two tutorial modules:

1. *Introduction to Proteomics Data.* This module will introduce the user to the format of both SELDI (Surface Enhanced Laser Desorption Ionization) and MALDI (Matrix Assisted Laser Desorption Ionization) data. This will include a short introduction to the mzXML format with references for further study. Also included will be suggestions on how to preprocess MALDI data for use with the Q5 algorithm. This module is optional depending on the user's familiarity with proteomics data.
2. *Introduction, Installation, and Usage of the Q5 R package.* This module will include references to downloading and installing the Q5 R-package, loading mzXML formatted data, appropriate preprocessing and formatting steps prior to Q5, basic usage of the Q5 algorithm, and interpretation of the results.

2.2 HANDS-ON TRAINING

This section of the training is an hour-long hands-on demonstration of the Q5 algorithm using example data provided by OHSU Adopters. Topics to be covered include: loading mzXML data for use in Q5, preprocessing MALDI data, running the Q5 algorithm, and interpretation of results. As pointed out above, this is optional.

3. AUDIENCE

Dartmouth (developers of the Q5 application) have indicated that the intended audience of Q5 is biostatisticians in various cancer centers. This end user of the Q5 program will use the tool to perform classification of proteomic spectra of serum samples collected within the cancer center and processed by an associated proteomics core. This end user is assumed to be familiar with the basics of the R Language interface and have a basic understanding of data reduction techniques.

4. DELIVERY

4.1 HANDS-ON TRAINING

This session is aimed at providing new users the necessary skills and knowledge to use Q5. Training sessions are scheduled on an as-needed basis. Delivery of this training requires a training facility outfitted with a computer lab with internet access, the R-language installed on each computer, and the ability to download Q5 and its associated dependent packages. Each training session can include up to 10 attendees from various institutions.

4.2 SELF-PACED LEARNING

This method of learning is suited for biostatisticians familiar with both the R Language and installing packages. Each tutorial will walk the user through a single component of the Q5 package in order to achieve a result. The tutorials are built such that completing them serially will result in understanding the full workflow needed to run the Q5 algorithm.

4.3 USER GUIDE

A user guide will be also be available online for Researchers, with appropriate references to the Developer-authored installation guide.