# 1. DOCUMENT CHANGE HISTORY

| Version Number | Date | Description |
|---|---|---|
| 1.0 | August 1, 2005 | Initial Draft |

# *Q5*

# Test Plan

# Version 1.0

**DRAFT**

# TABLE OF CONTENTS

# 2. INTRODUCTION

This Test Plan prescribes the scope, approach, resources, and schedule of the testing activities. It identifies the items being tested, features to be tested, testing tasks to be performed, personnel responsible for each task, and risks associated with this plan.

## 2.1 SCOPE

This document provides instruction and strategy for incorporating Software Testing practices and procedures into the Q5 project. This document demonstrates the application of testing on this project and provides guidelines for implementing procedures supporting this function.

The current Q5 application is an implementation of a probabilistic classification algorithm with demonstrated utility in classifying expression-dependent proteomic data from mass spectrometry of human serum. The Q5 project aims to extend the current implementation in the following ways i) implement the algorithm in an open source software environment (using the R statistical package and ii) ensure the implementation is compatible with caBIG. It was also agreed by the developer and adopter site that Q5 should be tested on higher resolution data (MALDI).

### 2.1.1 Identification

Q5 implemented in R

### 2.1.2 Document Overview

This Test Plan defines the strategies necessary to accomplish the testing activities associated with Q5. Testing procedural instructions are included in the Standard Operating Procedures (SOPs).

The remaining Test Plan sections are organized as follows:

- **Section 2: Strategy**: Describes the overall approach and techniques to be used during testing.
- **Section 3. Software Test Environment**: Describes the intended testing environment.
- **Section 4. Test Identification**: Identifies and describes each of the tests.
- **Section 5. Test Schedules**: Contains or references the schedules for conducting testing.
- **Section 6. Requirements Traceability:** Identifies traceability from this Test Plan to the Requirements.
- **Section 7. Risks**: Identifies the risks associated with the Test Plan.
- **Section 8. Notes:** Contains general information that aids in the understanding of this document.

## 2.2 RESOURCES

Shannon McWeeney: PI and manager of Oregon health and Science University Q5 adopter site

Ted Laderas: Test Manager

Ranjani Ramakrishnan: software/statistical algorithm tester

Cory Bystrom: Proteomics SME at Oregon Health and Science University

Paul Courtney: Manager of Development at Dartmouth

## 2.3 REFERENCED DOCUMENTS

For additional project specific information, refer to the following documents:

- Q5 Vision and Scope Document (http://cabigcvs.nci.nih.gov/viewcvs/viewcvs.cgi/qfive/caBIG_Q5_Vision_Scope_Draft.doc )
- Q5 requirements and Specifications document (http://cabigcvs.nci.nih.gov/viewcvs/viewcvs.cgi/qfive/caBIG_Requirements_and_Specification_Q5Project_DRAFT.doc)
- Q5 Use Case Documents (http://cabigcvs.nci.nih.gov/viewcvs/viewcvs.cgi/qfive/ )
- Meeting Notes August 1, 2005

# 3. SOFTWARE TEST STRATEGY

## 3.1 OBJECTIVES

We propose to carry out white box (structural testing), validating the implementation and limited performance and user acceptance testing.

## 3.2 APPROACH

**White-Box (Component) Testing of Q5 components**

We will test the individual components (PCA and LDA) of the Q5 algorithm on well-characterized datasets (such as Fisher's Iris dataset). This will allow determination that the individual algorithms are correctly implemented. We will then compare the output of the individual components from the original SELDI data sets to ensure that the port to R was successful by comparing it with the output for these components with the original implementation (Matlab).

**Validation Testing**

We will validate the output (i.e. spectral classification and cross-validation statistics) of the Q5 Algorithm on two biological datasets. Validation will be performed on higher resolution MALDI data as agreed by the developer and adopter.

Vanderbilt dataset – this is a set of serum samples from lung cancer and normal patients processed as MALDI spectra.

OHSU Cancer Institute dataset – this is a prostate cancer dataset that OHSU will produce and will consist of three kinds of samples: patients diagnosed with prostate cancer, patients with elevated PSA levels but with benign hyperplasia, and normal patients. This dataset will also consist of MALDI spectra.

Because the Q5 algorithm requires that the number of features must be smaller than the number of samples submitted, we will utilize the statistical processing routines available as part of the caBIG RProteomics project (for which OHSU is an adopter site) in order to perform this dimensionality reduction step, rather than using the raw spectra. This approach is based on our conversations with the developer August 1$^{st}$ and the limitations of the developer's SOW at this time. We will document which statistical processing routines were used.

**Performance Testing**

We will measure CPU usage, memory usage, and execution time for both datasets and report these results on Windows, Solaris, and Linux machines.

**Usability Testing**

We will perform usability testing of Q5 with a number of users recruited from the OHSU Cancer Institute and report their general impressions and suggestions for making Q5 useful.

## 3.3 DESCRIPTION OF FUNCTIONALITY

Q5 is a probabilistic classification algorithm implemented using R. It accepts as input SELDI TOF and high resolution MALDI TOF data (if the dimensions of the data are reduced sufficiently) and outputs a classifier.

## 3.4 SPECIFIC EXCLUSIONS

Integration testing for the grid – This is not specified by the developer in the Q5 Requirements and Specifications document.

Functional testing – The adopters are not completely blind to the Q5 code.

Unit testing – This will be performed by the developers

## 3.5 DEPENDENCIES & ASSUMPTIONS

Delivery of software by the developers.

## 3.6 GENERAL CRITERIA FOR SUCCESS

Correctness of implementation validated.

All tests have been executed and the outputs documented, resolved, verified, or designated for future releases.

Report for users on speed and resource utilization for different platforms.

### 3.6.1 Readiness Criteria

Will be communicated by the developers.

### 3.6.2 Pass/Fail Criteria

This is dependent upon the output format being finalized by Dartmouth (e.g., the final Requirements and Specifications document).

### 3.6.3 Completion Criteria

The criteria for completion of the testing procedures is that the system produces the output (described in the Requirements and Specifications document) within expected performance requirements. Testing is considered completed when:

- The assigned test scripts have been executed.

- Defects and discrepancies are documented, resolved, verified, or designated as future changes.

### 3.6.4 Acceptance Criteria

This is dependent upon the output format being finalized by Dartmouth (e.g., the final Requirements and Specifications document).

# 4. SOFTWARE TEST ENVIRONMENT

This section describes the software test environment at each intended test site.

## 4.1 OREGON HEALTH AND SCIENCE UNIVERSITY

### 4.1.1 Software Items

Microsoft Windows XP
R 2.1.1
Solaris

### 4.1.2 Hardware and Firmware Items

Solaris Server Configuration (Murdock 1)
Windows machines
Pentium$^{R}$ 4 3.6 GHz, 1 GB RAM

### 4.1.3 Other Materials

The other major components are described below.

Vanderbilt data

OHSU data

### 4.1.4 Participating Organizations

The testing group consists of the project's Test Manager (TM), and the Tester(s). The groups listed below are responsible for the respective types of testing:

| **Tester** | **Responsibility** | **Qualification** |
|---|---|---|
| **Shannon McWeeney** | **Oversight** | **PI and Project manager** |
| **Ted Laderas** | **Testing** | **Test Manager** |
| **Ranjani Ramakrishnan** | **Testing** | **Tester** |

# 5. TEST SCHEDULES

Our schedule is determined by the final specification of the output format by Dartmouth and delivery of the final Q5 code.

## 5.1 TIME FRAMES FOR TESTING

The Test Manager will coordinate with the Project Manager and add the planned testing activities to the master project schedule. Refer to the project SDP and schedule for additional information.

# 6. RISKS

Handling the high resolution MALDI-TOF data is dependent on the dimensionality of the data being reduced sufficiently to allow its being used as input to Q5. As the SOW for the developer prevents actual extensions to the algorithm, this will be dependent upon the success of the routines from Rproteomics.

# 7. [NOTES]