



**CALIFORNIA STATE SCIENCE FAIR
2010 PROJECT SUMMARY**

Name(s) Tony Ho; Ritik Malhotra	Project Number S1604
--	---------------------------------------

Project Title
Novel Genetics-Based Early Disease Detection Using Ontology-Driven Microarray Semantics with Optimized Mass Spectrometry

Abstract

Objectives/Goals
Doctors currently use symptoms to determine whether certain diseases are present, which is prone to error and subjectivity. Genetics allow for earlier and accurate diagnosis of diseases. Microarray experiments hold data describing genes that cause specific diseases, but current databases are disorganized and lack semantic enforcement. Mass spectrometry techniques to analyze tissue samples exist as a viable means to use data from microarrays to find diseases, but the current mass spectrometers aren't sensitive enough to find the diseases at low concentrations, before they become lethal. The purpose of this project is to identify key biomarkers for diseases by detecting the genetic changes caused by diseases. First, an efficient database to query microarray experiments without inaccurate data is needed. Second, sensitivity in mass spectrometers needs to be improved in order to facilitate the early detection of low abundance biomarkers.

Methods/Materials
A cluster of 8 quad-core 64 bit servers and an Agilent G6460 triple-quadrupole mass spectrometer with Agilent MassHunter software were used. First, a MySQL database using text mining techniques for the microarray experiment data was designed. Second, the inaccurate data was removed from the database by developing two novel algorithms: KNN-Delta and Semantic Outlier Factor. Third, thermal gradient focusing technology in the mass spectrometer for ion confinement was implemented. Fourth, conditions inside the mass spectrometer were optimized for efficient ion transmission. Fifth, the ion scattering within the electrospray ionization unit of the mass spectrometer was analyzed to detect areas of high densities. Finally, an inlet capillary to capture ions in the high density areas was designed.

Results
The microarray database was successfully created and was able to automatically annotate over 400,000 experiments using common semantics. The KNN-Delta and Semantic Outlier Factor algorithms were able to increase the accuracy of the database, removing 40,000 inaccurately annotated experiments. The mass spectrometry efficiencies resulted in a signal-to-charge ratio increase from 1000:1 to 6800:1, detecting samples as low as 150 femtomoles.

Conclusions/Discussion
Based on the successful implementation of the microarray database and the sensitivity improvements of the mass spectrometer, it is possible to detect diseases at small concentrations, before they become lethal.

Summary Statement
By developing novel algorithms for a microarray database and improving the sensitivities of a mass spectrometers, we are able to facilitate the early detection of diseases, before they become lethal.

Help Received
First, we want to thank Dr. Paul Momoh from Agilent Technologies and Dr. David Ruau from Stanford University for their mentorship. Next, we want to thank our parents, peers, Mrs. Amanda Alonzo, Dr. Roy Rocklin, and the rest of the Lynbrook administration for their constant support in our scientific ventures.