DecisionQ MEMORANDUM

To: Dr. Gregory Downing
Personalized Healthcare Initiative

Date: January 2, 2007

From: John Eberhardt

Re: Response to RFI for Personalized Medicine Technologies

Summary

DecisionQ is a leading company in the field of data mining and predictive analysis. Based on research developed at the Stanford Aerospace Robotics Lab, DecisionQ has applied advanced heuristic learning techniques and complex information structures to common problems in life sciences, healthcare, consumer products, financial services, and national security.

DecisionQ software uses Bayesian networks to represent relationships between variables, even if the relationships involve uncertainty, unpredictability or imprecision. Historically, automatically determining Bayesian networks was very difficult. DecisionQ has broken through this barrier, using machine learning technology to enable the automated creation of Bayesian Networks for huge data sets quickly, using inexpensive computer equipment.

Machine learning is a field of computer science that uses intelligent algorithms to allow a computer to mimic the process of human learning. The DecisionQ machine learning algorithm allows the computer to learn dynamically from the data that resides in the data warehouse. The machine learning algorithms automatically detect and promote significant relationships between variables, without the need for human interaction. This allows for the processing of vast amounts of complex data quickly and easily.

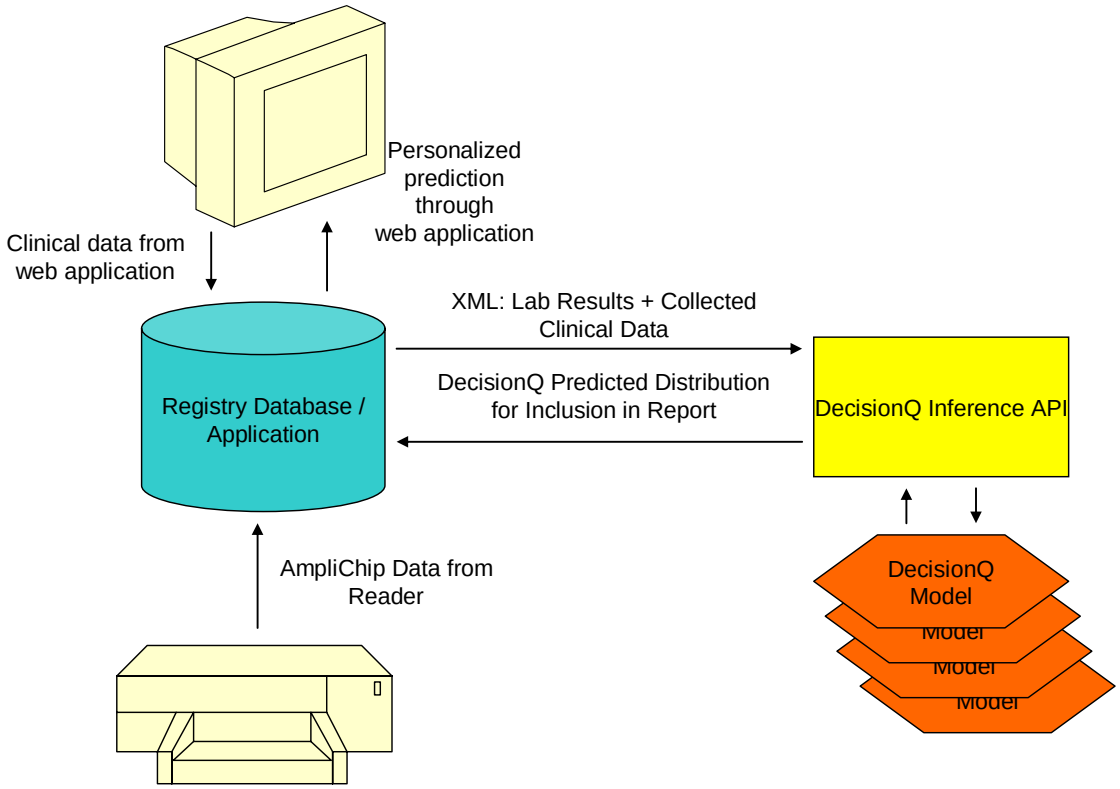
DecisionQ's scientists and employees are widely acknowledged as leaders in this field and are frequent speakers at data mining and knowledge discovery conferences. Some selected employee publications in life sciences include *Using Microarrays for Molecular Diagnostics: An Application to Identify Tumor Site of Origin*, Journal of Molecular Diagnostics November 2004, Vol. 6, No. 4; *Gene Expression Array-Based Diagnostics for Identifying Tumor Site of Origin*, Proceedings of the 2004 Annual Meeting of the American Association of Clinical Chemists (AACC), Los Angeles, CA; *Application of multivariate probabilistic (Bayesian) networks to the interpretation of diagnostic assays* published in the proceedings of the 2005 AACC Oak Ridge conference in Baltimore; and

As a company, DecisionQ is committed to supporting personalized medicine with technologies that allow for the rapid comparison of individual patient genotype, phenotype, and diagnostic information with broad population information to develop i) a very sensitive and specific patient diagnosis, ii) patient specific risk-stratification and prognosis, and iii) probabilistic scoring of likely response to different treatment protocols, based on broad population evidence.

DecisionQ Corporation will deploy their integrated predictive modeling application combined with broad population data to allow the clinician to interpret diagnostic results in the context of large population evidence quickly and automatically, with detailed risk statification and quality estimates. DecisionQ calls these models Population Evidence Models©.

Population Evidence Models© Overview

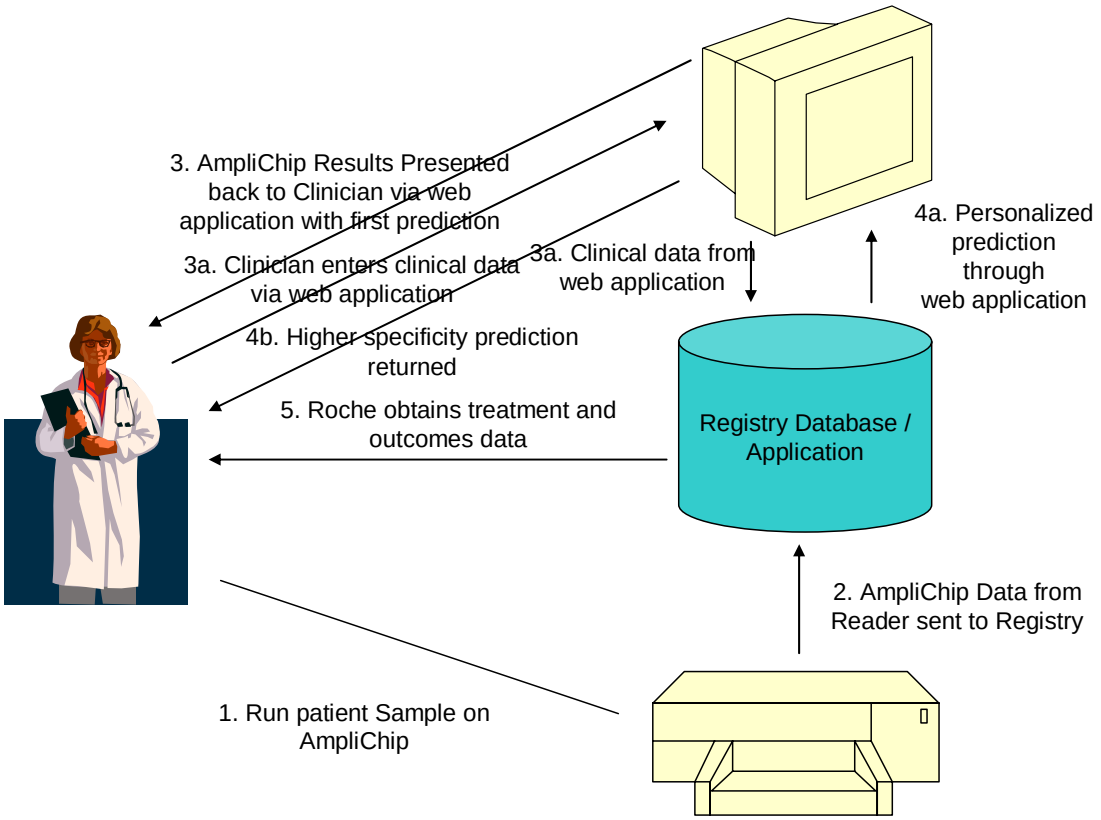
Population Evidence Models have a specific architecture, consisting of a data repository/registry, a reference model or models, a web interface that allows clinicians to interact with the models, a messaging architecture for moving information between components, and an overall process for system update, management, audit, and QC.



Use Case

In order to better understand how the user will interact with the system, the following use case has been developed:

1) In order to assess potential therapies for John Doe, Dr. Wilson runs a diagnostic – in this case a p450 AmpliChip
2) When the analyzer completes the assay, the chip values are returned in a non-specific format to the registry for analysis
3) The patient values are then presented back to the clinician along with baseline model prediction for therapies and outcomes along with frequency estimates and quality statistics
   a) At this point, the clinician has the option of entering additional patient data to recalculate probabilities and frequency estimates and error/quality statistics



Population Evidence Model Components

Data Repository/Registry

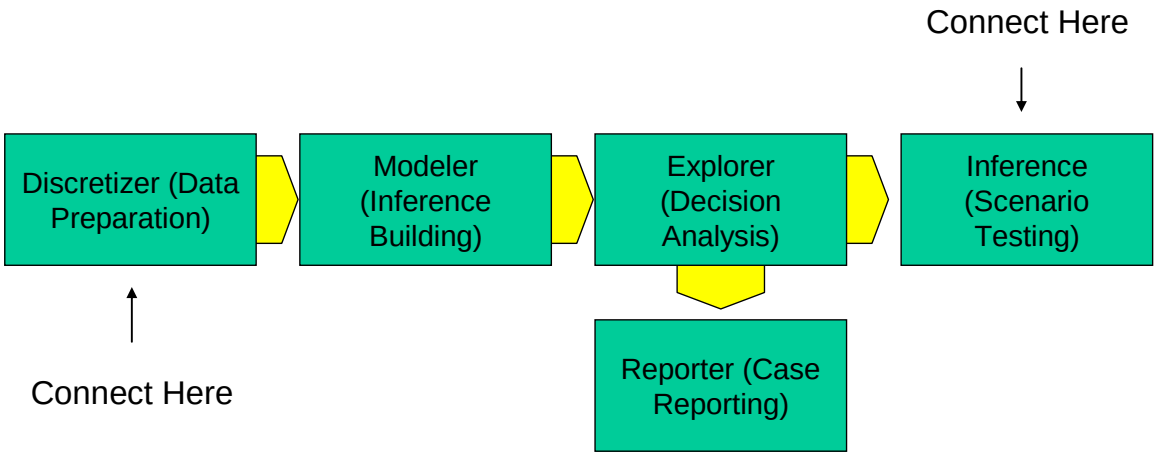
The data repository needs to support reference modeling, cross-validation, and prediction for the initial product launch and will need to sustain ongoing model updates, validation, and prediction for a substantial period of time. Thus, the data repository must meet several criteria:

- Thorough schema design – even if there is no intention of collecting a certain datum in the near future, we must anticipate this in the design of the schema
- Flattenable schema for modeling and prediction – standard queries must be developed that can be used to flatten the schema for analysis and prediction without re-writing the queries

- Scalability – if the project is successful, the repository must be able to store substantial amounts of data
- Population with initial focus criteria
- Identification of initial focus criteria that are "collectable" given physician workflow

Reference Model

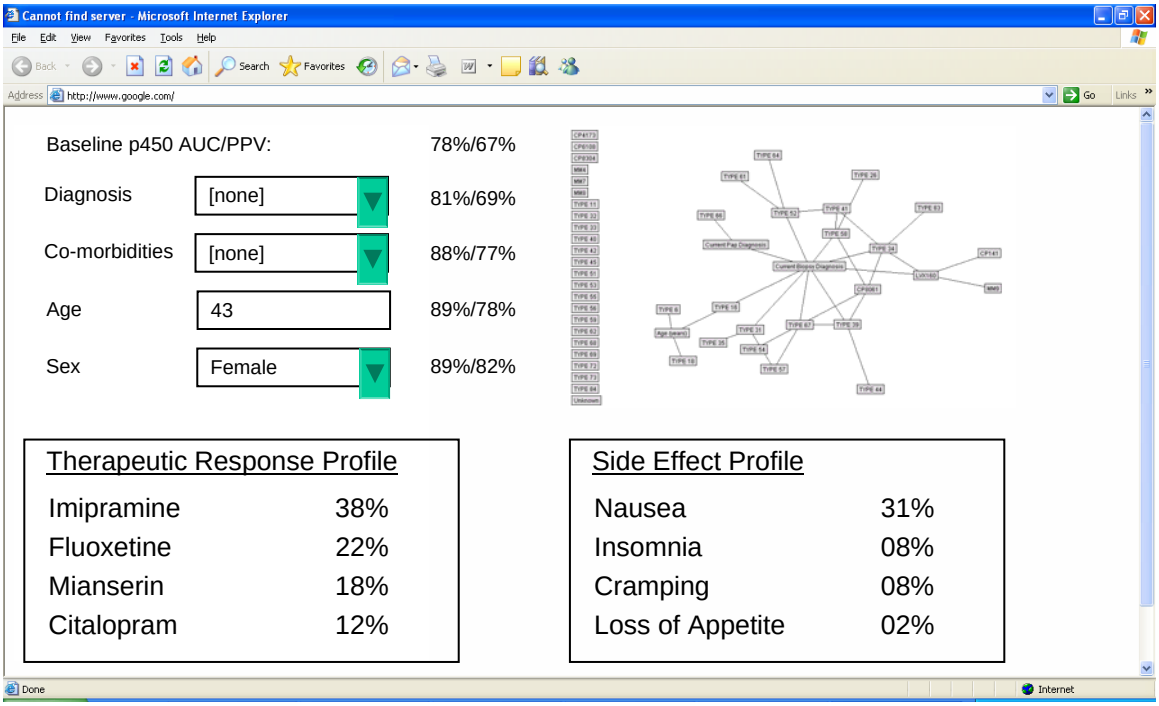
The reference model(s) are DecisionQ BN5 format models and will use the genotype and phenotype criteria to develop case specific predictions for each patient. These models will utilize the training table in the database to construct evidence-based predictive Bayesian (probabilistic) networks. Using this data, the models will be automatically updated using DecisionQ unsupervised Bayesian networks.



As each model is created, a validation and verification documentation set will be created that will consist of a data file, model and settings file, a modeling log, and cross-validation statistics. As reference models are updated for new results, the new reference model will be inserted into the system as an XML component and the old reference model will be archived.

Web Interface

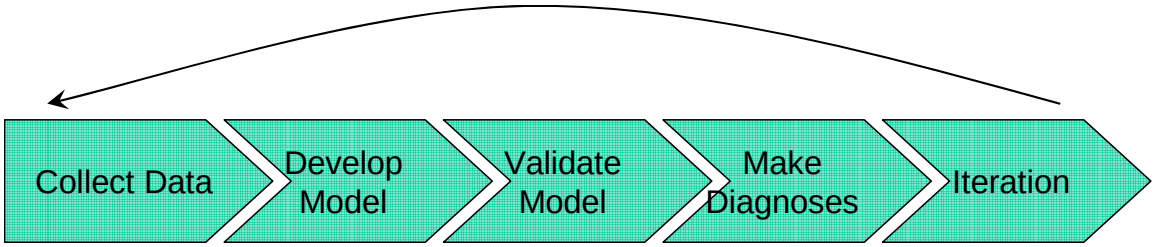
Clinicians will access the system through a web interface that will allow them to receive personalized predictions, input relevant case information to improve the specificity of the prediction, and receive an updated prediction. The example below presents one option of how this interface might look. Once the clinician has input the relevant data and received their updated prediction, this report can be saved in HTML format, output to a CSV file, or written to the LIS or EMT systems through a middleware product.



Along with a web prediction interface, we can provide an optional downloadable viewer to allow clinicians to manipulate models directly if they so desire.

Control Process

It is important to note that this system/product is, by definition, not static. It requires a process of constant update, validation, and improvement. As new data is collected, models will be updated and QC/QA documented. In general, the graphic below represents the iterative, ongoing process for product deployment.



Each iteration of this process will produce new reference models and accompanying validation and verification documentation. There will need to be assigned personnel and permissions to support this process.

Supporting Documentation

1) FasterAnalytics White Paper
2) IVD Technology Article on Modeling Renal Disease
3) San Antonio Breast Cancer Symposium Poster on Modeling Breast Cancer Recurrence