QUANTITATIVE MEDICINE

Transforming Drug Discovery™

Technical Discussion

Joshua Kangas PhD

Founder & Chief Science Officer

Technology Evaluation Consortium

Cambridge Healthtech Associates

October 23, 2013

QUANTITATIVE MEDICINE
Transforming Drug Discovery™

This Presentation Contains Proprietary and Confidential Information

It may not be shared with any third party or be used for any purpose other than to evaluate a potential business relationship and/or perform a TEC study with Quantitative Medicine.

Open Question

How can we more effectively utilize ***all*** publicly available screening results improve prediction accuracy?

QUANTITATIVE MEDICINE
Transforming Drug Discovery™

New prediction methods utilizing all historical data

- For each assay in the historical data, learn a QSAR model using all available data (10-200,000 discrete observations)
- These models only predict what was measured in the experiment (not necessarily hepatotoxicity).
- Make predictions for all compounds “observed” and to be predicted.
- These predictions form the basis of a feature matrix describing each compound.
- Learn a model to predict hepatotoxicity from the resulting feature matrix where observations are available

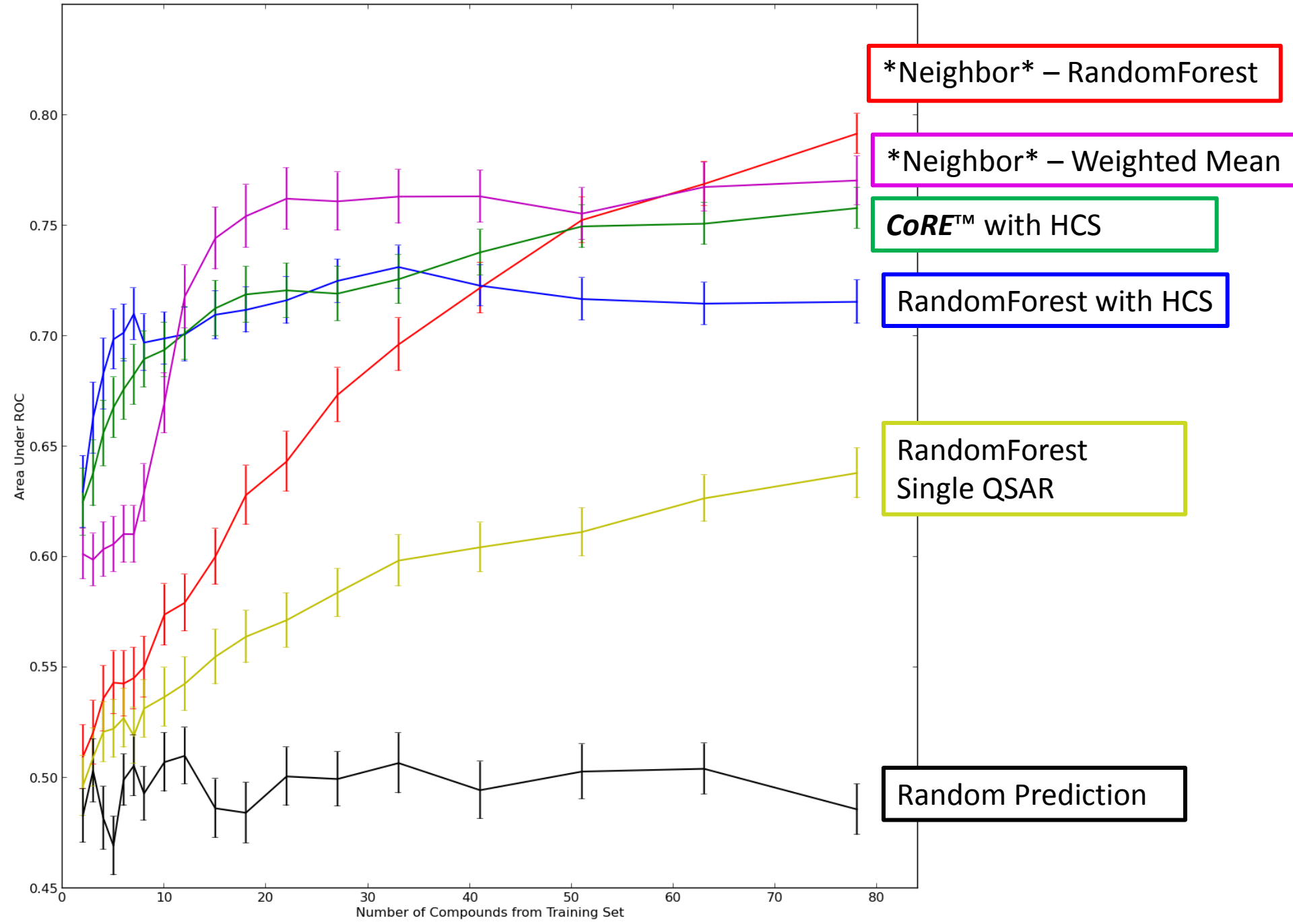
New prediction methods utilizing all historical data

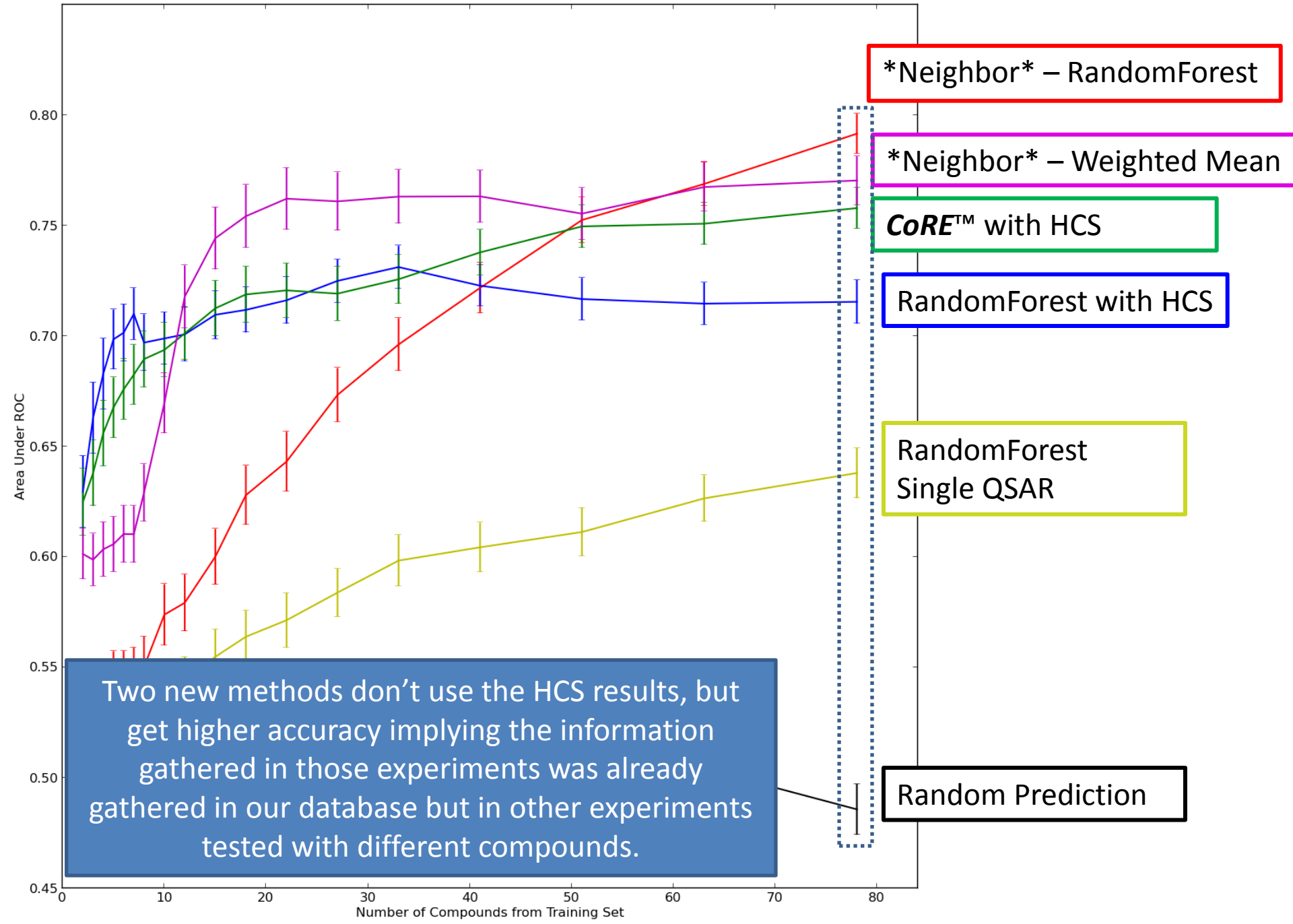
- Random Forest – use random forest to learn a model to predict hepatotoxicity from prediction based features
- Weighted mean – for each assay, determine the correlation between model predictions and observations. Weight that assays predictions for held out set by –log(p) where p is the probability of seeing a higher correlation in a permutation test.

QUANTITATIVE MEDICINE
Transforming Drug Discovery™

Prediction Simulations

- Selection Method: Random
- Batch size: 20% of what had been previously observed (cuts down on run time in generally uninteresting phases when there are a lot of observations)


Neighbor – RandomForest
Neighbor – Weighted Mean
CoRE™ with HCS
RandomForest with HCS
RandomForest
Single QSAR
Random Prediction
Area Under ROC
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0
10
20
30
40
50
60
70
80
Number of Compounds from Training Set


Neighbor – RandomForest
Neighbor – Weighted Mean
CoRE™ with HCS
RandomForest with HCS
RandomForest
Single QSAR
Random Prediction
Two new methods don't use the HCS results, but get higher accuracy implying the information gathered in those experiments was already gathered in our database but in other experiments tested with different compounds.
Area Under ROC
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0
10
20
30
40
50
60
70
80
Number of Compounds from Training Set

Other considerations

- Experiment direction
- Types of model predictions
  - Neighbor approach with RF can use any types of predictions (discrete or continuous)
  - Weighted mean neighbor approach can only use two-class discrete predictions unless predictions are scaled from 0 – 1.
- Next steps?

Quantitative Medicine

Transforming Drug Discovery™

Joshua D. Kangas PhD
josh.kangas@QtMed.com
Founder &
Chief Science Officer

John R. King
john.king@QtMed.com
President & CEO

www.QtMed.com