QUANTITATIVE MEDICINE

TRANSFORMING DRUG DISCOVERY

Technology Evaluation Consortium
Cambridge Healthtech Associates

September 13, 2013

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.

Agenda

- ToxCast Study Design
- Simulation Results
- Prospective Uses
- Next Study

ToxCast Study Design

Purpose: Develop a Model to Actively Learn and Accurately Predict All Observations in ToxCast

ToxCast Data

- Total Assays: 1197
  - ACEA – change in cell growth kinetics
  - AttaGene – reporter gene assay
  - BioSeek - ELISA
  - Cellumen – High Content Screeening
  - CellzDirect – change in expression levels
  - GenTronix – *in vitro* measure of growth arrest
  - NCGC - reporter gene assay
  - NovaScreen – enzyme inhibition assays
  - Solidus – Microarray?
  - ToxRefDB – Many toxicity measures

ToxCast Data

- Compounds: 309
- Experimental results:
  - Total possible results: 369,873
  - Available results: 321,033 (87% coverage)
  - Continuous measures from database
  - Discrete inferred from Continuous measures

How do we determine what is Ground Truth?

- Threshold data based on ToxCast results

CASRN	NAME	CLM_CellLoss_24hr	CLM_Hepat_Apoptosis_24hr	CLM_Hepat_DNADamage_24hr
50594-66-6	Acifluorfen	1000000	1000000	1000000
15972-60-8	Alachlor	61.46	1000000	1000000
116-06-3	Aldicarb	1000000	1000000	1000000
834-12-8	Ametryn	1000000	1000000	1000000
33089-61-1	Amitraz	123.9	144	NA
101-05-3	Anilazine	1000000	1000000	1000000
3337-71-1	Asulam	1000000	1000000	1000000
1912-24-9	Atrazine	1000000	1000000	1000000
35575-96-3	Azamethiphos	102.4	1000000	52.1
131860-33-8	Azoxystrobin	1000000	1000000	1000000

How do we determine what is Ground Truth?

- Threshold data based on ToxCast results

CASRN	NAME	CLM_CellLoss_24hr	CLM_Hepat_Apoptosis_24hr	CLM_Hepat_DNADamage_24hr
50594-66-6	Acifluorfen	1000000	1000000	1000000
15972-60-8	Alachlor	61.46	1000000	1000000
116-06-3	Aldicarb	1000000	1000000	1000000
834-12-8	Ametryn	1000000	1000000	1000000
33089-61-1	Amitraz	123.9	144	NA
101-05-3	Anilazine	1000000	1000000	1000000
3337-71-1	Asulam	1000000	1000000	1000000
1912-24-9	Atrazine	1000000	1000000	1000000
35575-96-3	Azamethiphos	102.4	1000000	52.1
131860-33-8	Azoxystrobin	1000000	1000000	1000000

Receiver Operating Characteristic

- What are we predicting?
  - Assay results for all “unobserved” experiments in ToxCast
- How do we measure accuracy of those predictions?
  - Calculate a ROC curve by sweeping across prediction thresholds and comparing to hidden ground truth in ToxCast

Compound Diversity-Based Selection

Purpose: Develop a Model to Actively Learn and Accurately Predict All Observations in ToxCast

Standard Approach (single pass):

Selection method: Cluster all compounds by their fingerprints into K clusters. Choose one compound from each cluster.

Prediction method: Learn a QSAR model using Lasso regression or RandomForest.

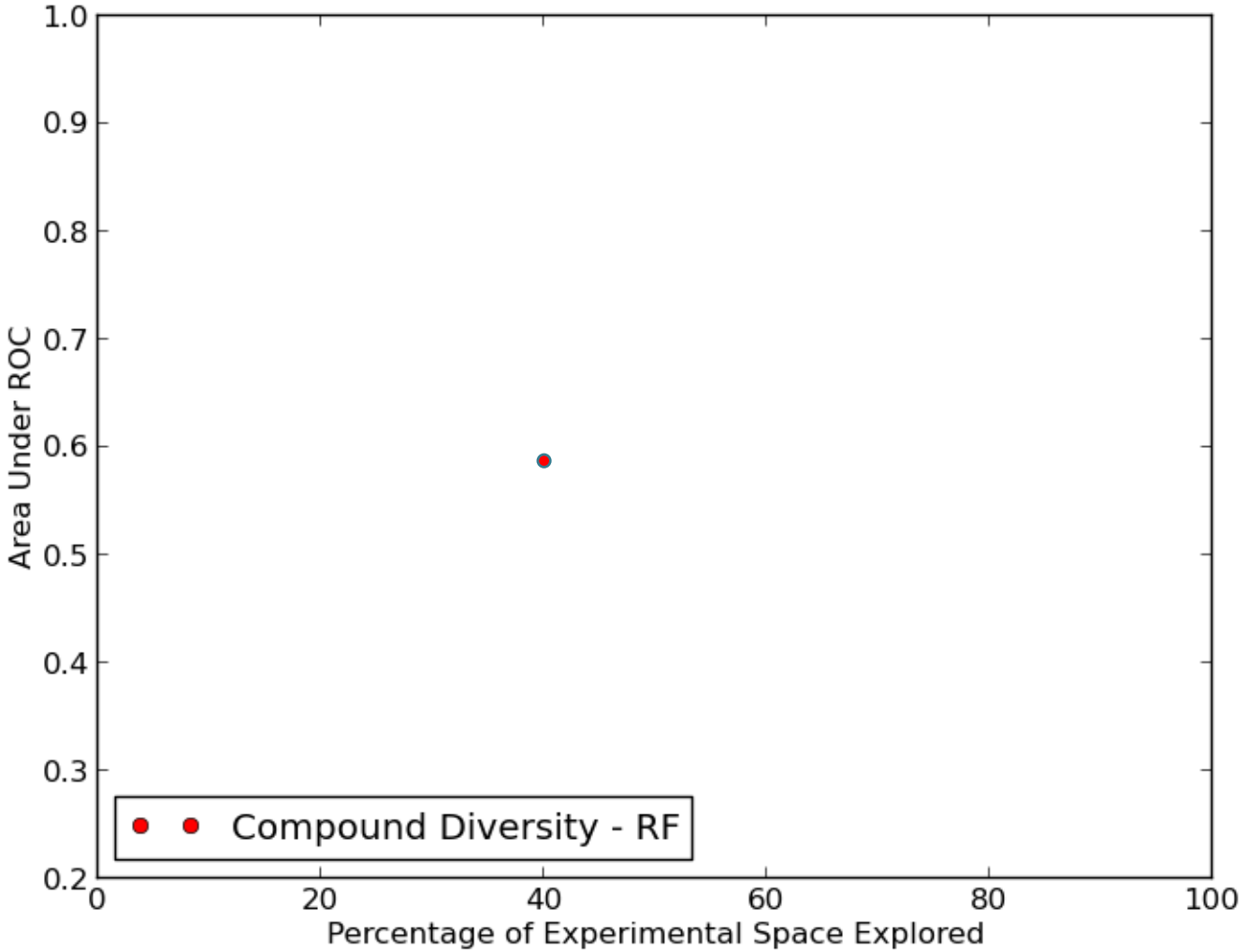
AFRS-Based Experiment Selection

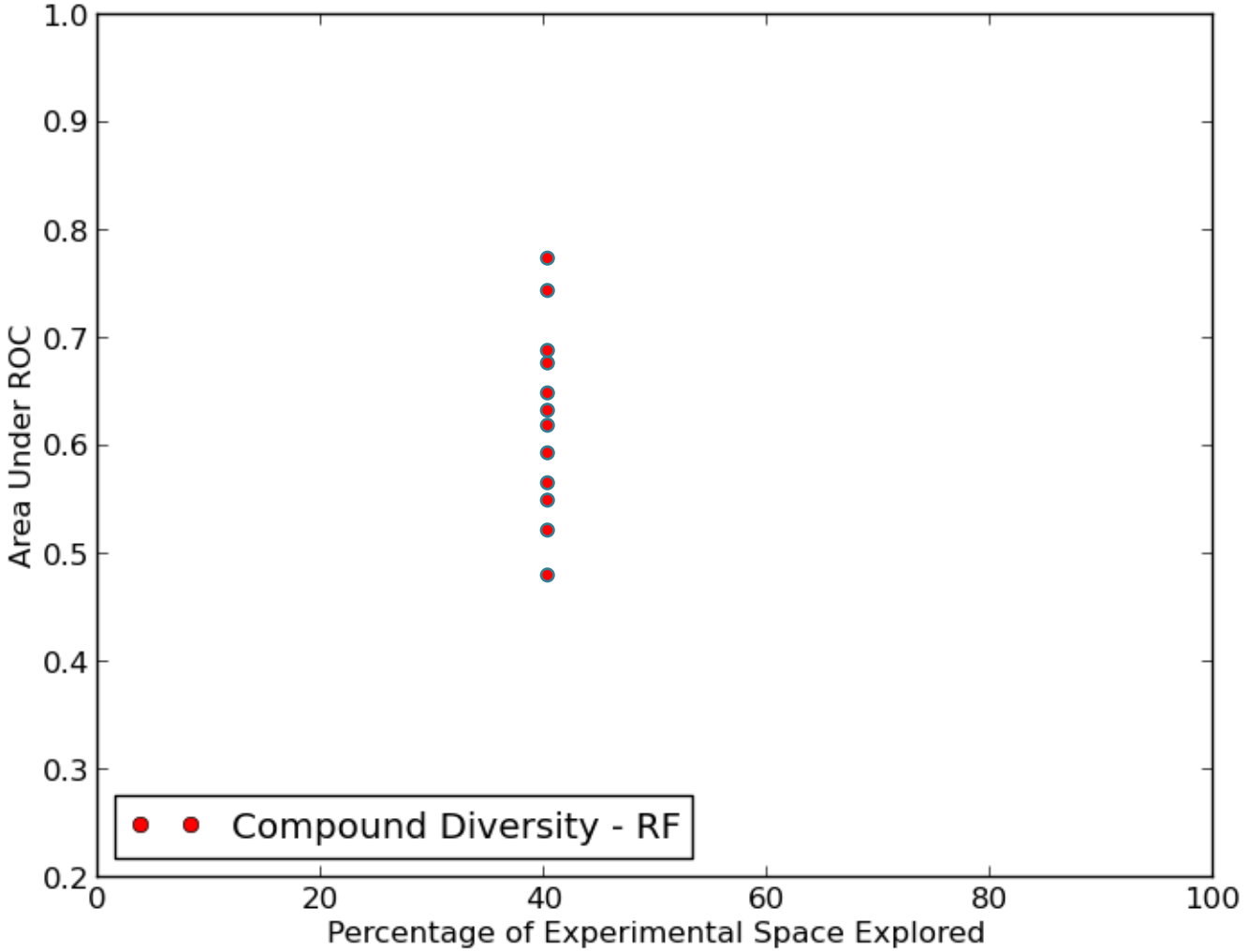
Purpose: Develop a Model to Actively Learn and Accurately Predict All Observations in ToxCast

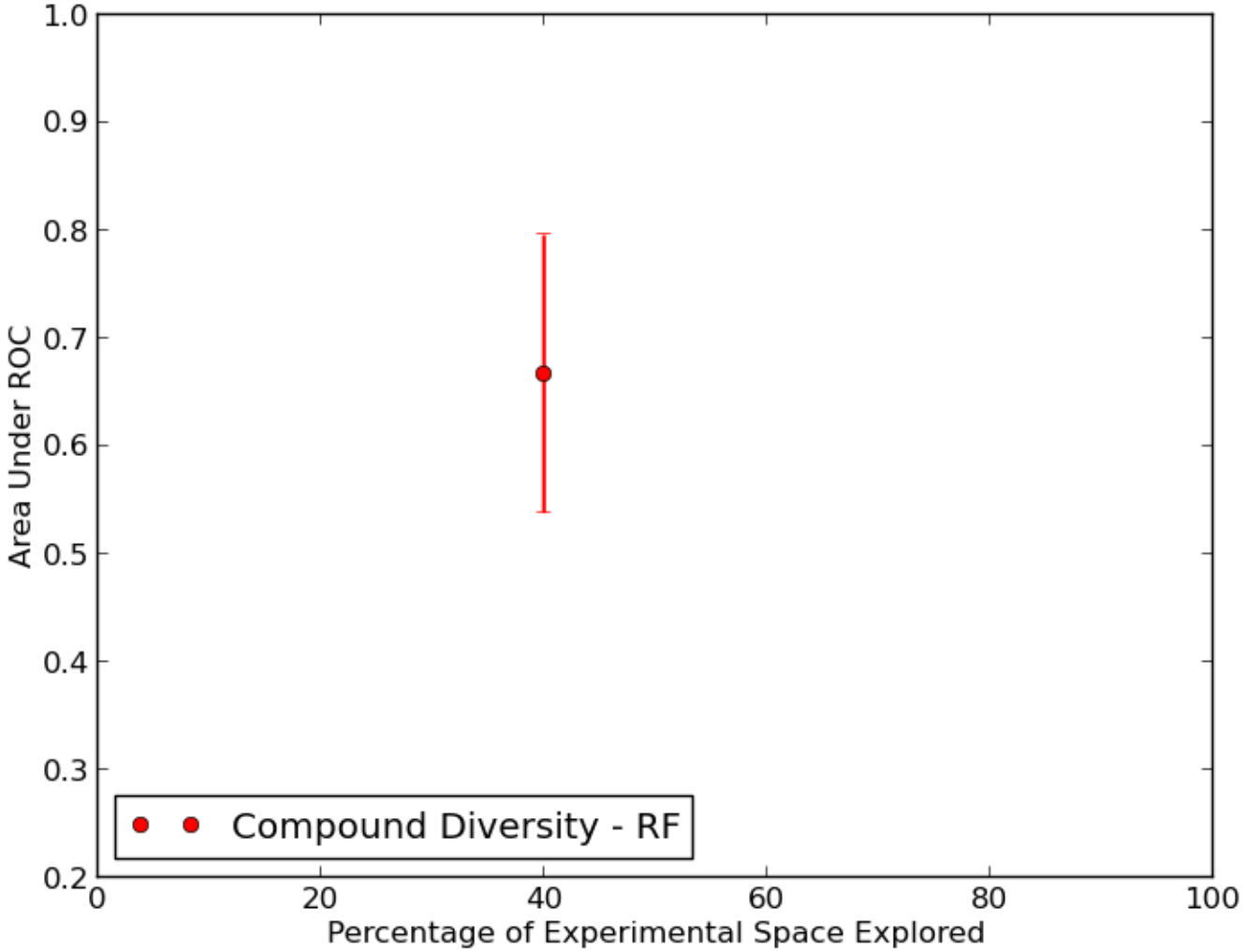
AFRS Approach (iterative):

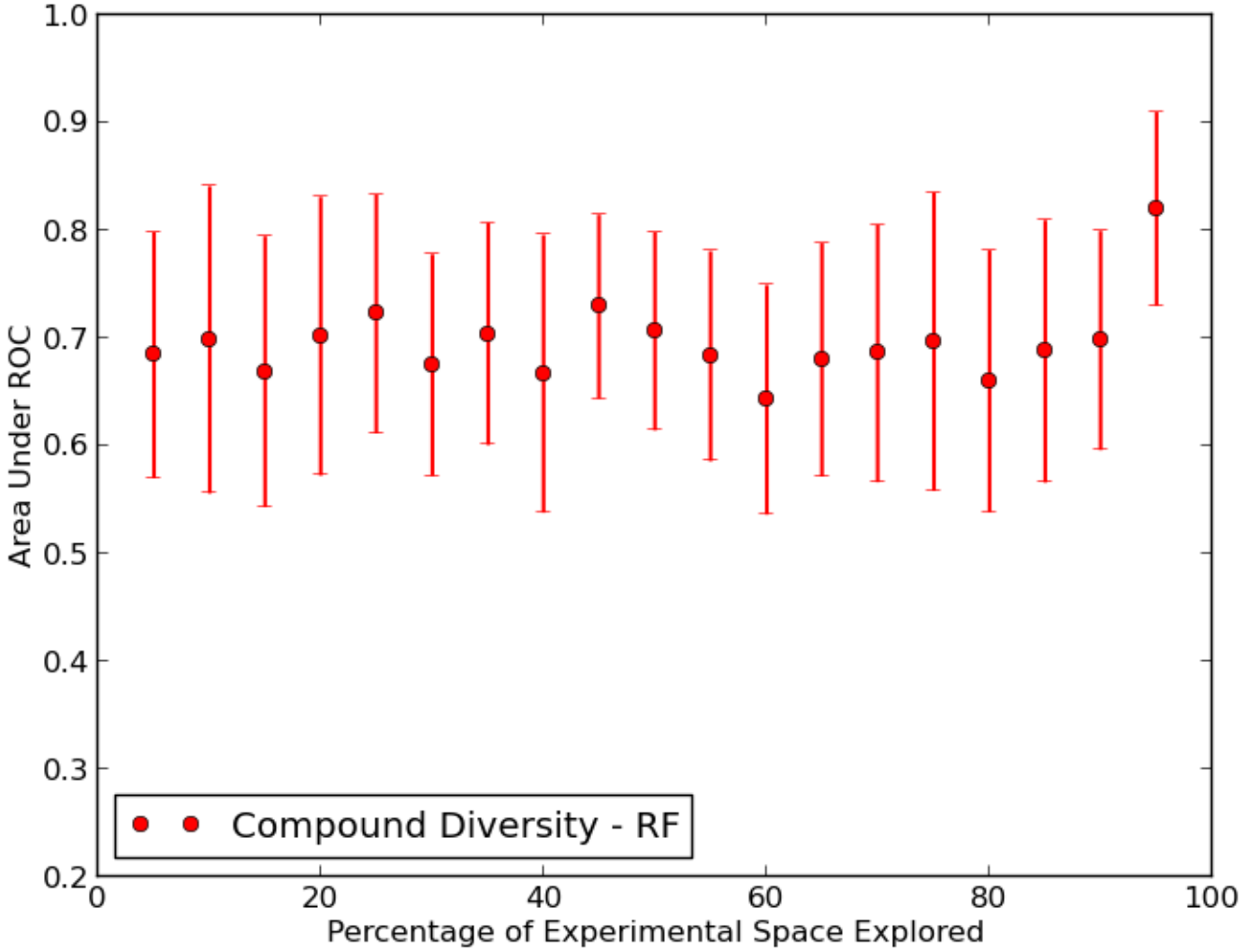
AFRS Selection Method: Choose batches of 1% of experiments believed to be the most informative.

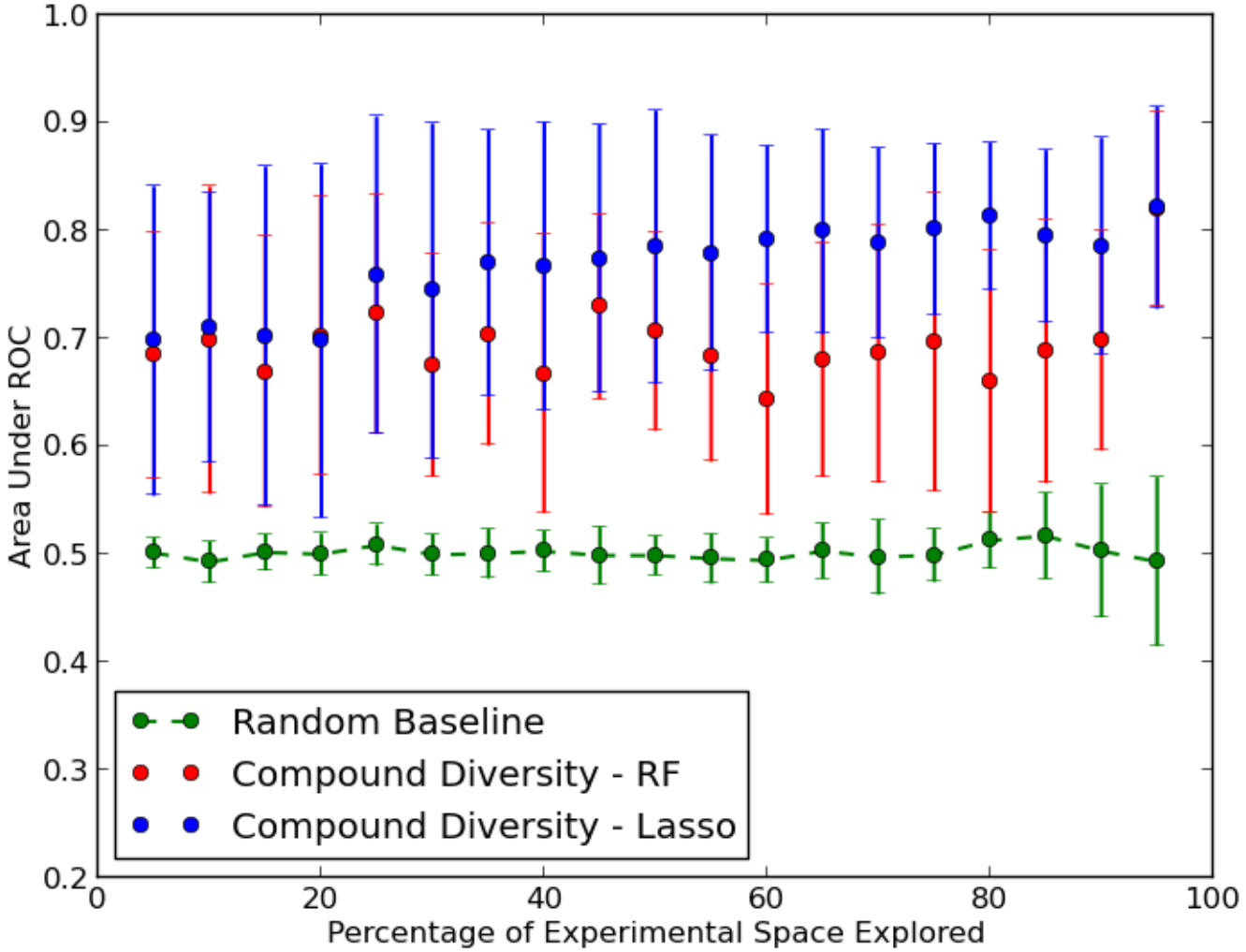
AFRS Prediction Method: Learn a model using the AFRS for to predict the results of all untested experiments.

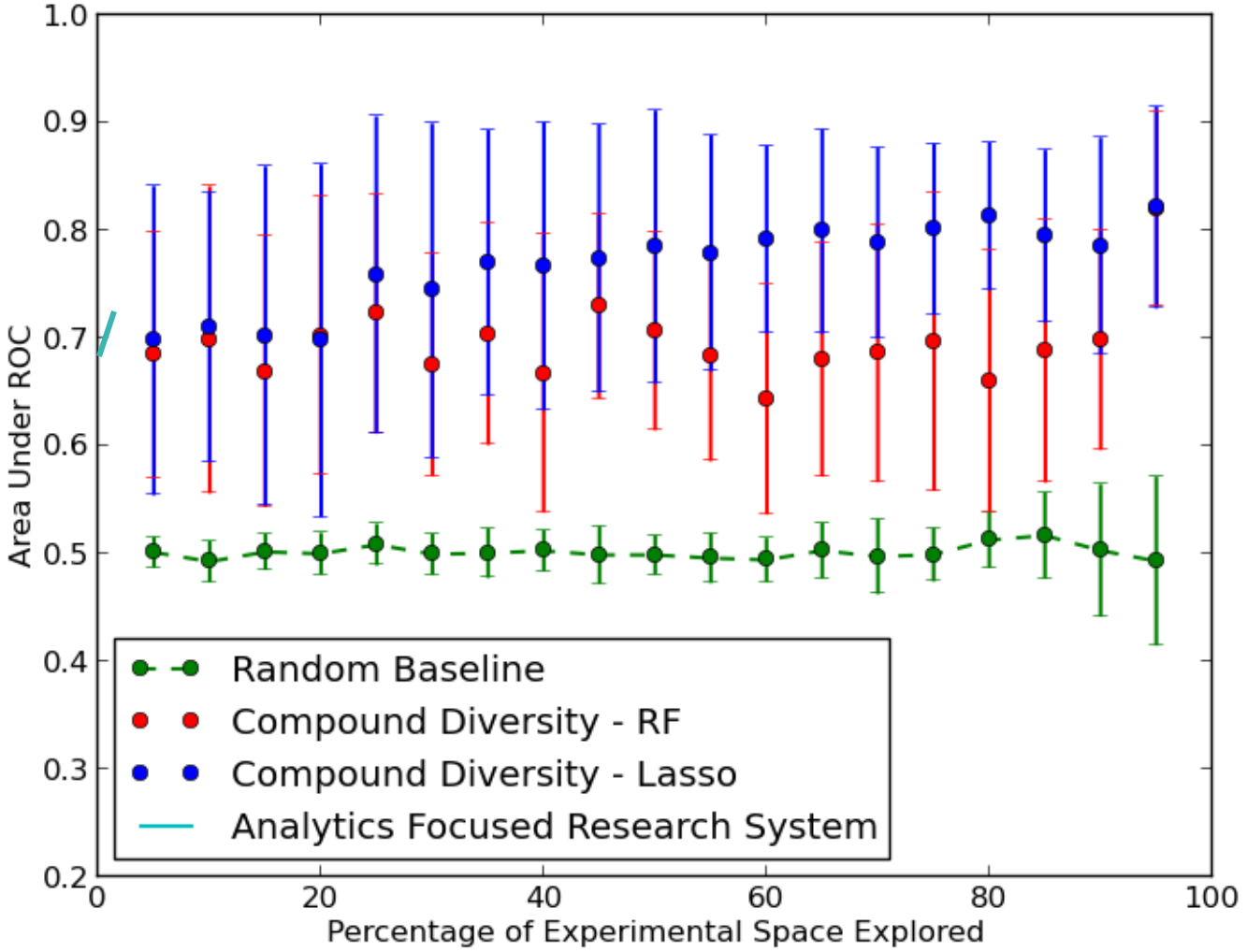

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Area Under ROC
0
20
40
60
80
100
Percentage of Experimental Space Explored
Compound Diversity - RF


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Area Under ROC
0
20
40
60
80
100
Percentage of Experimental Space Explored
Compound Diversity - RF

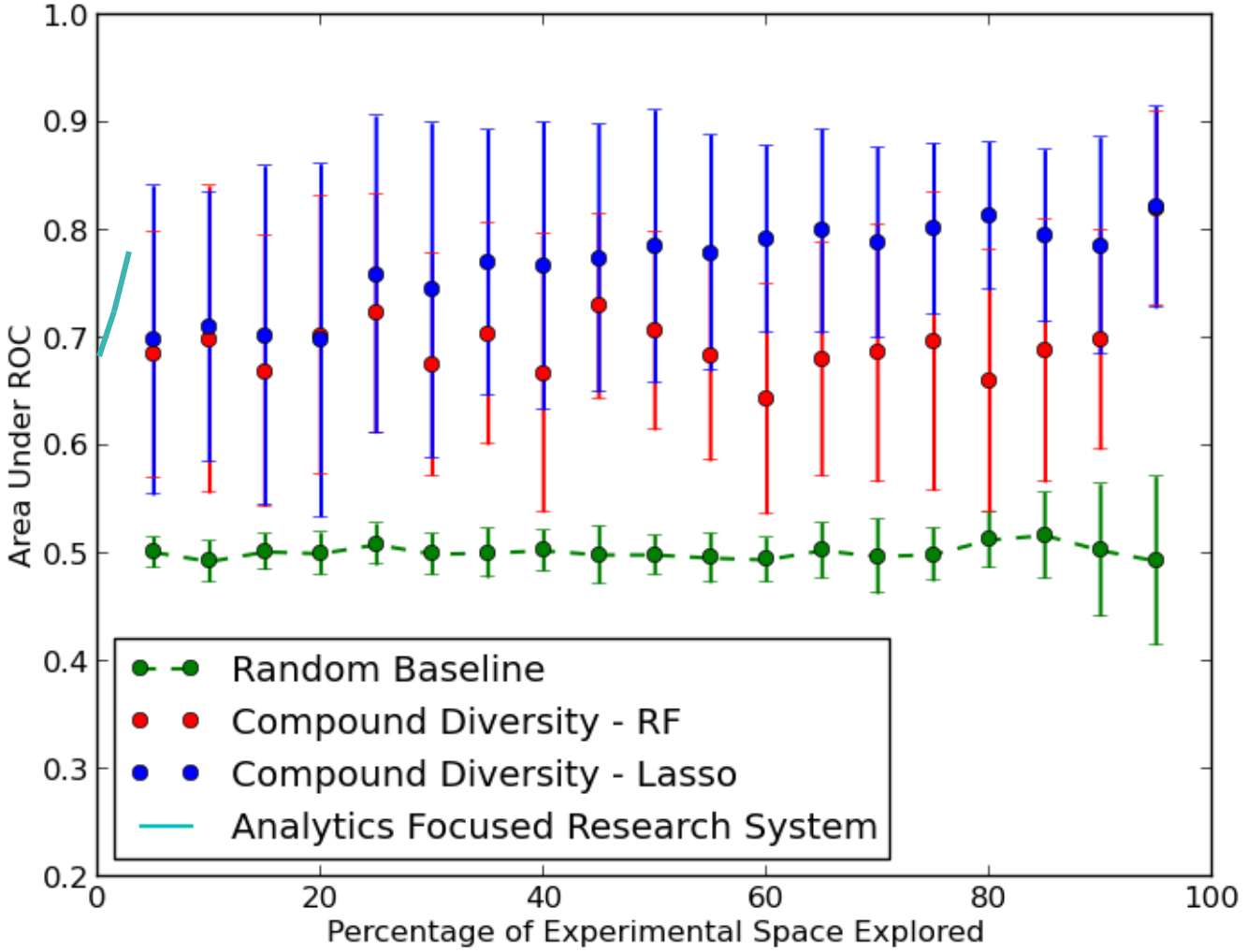

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Area Under ROC
0
20
40
60
80
100
Percentage of Experimental Space Explored
Compound Diversity - RF


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Area Under ROC
0
20
40
60
80
100
Percentage of Experimental Space Explored
Compound Diversity - RF

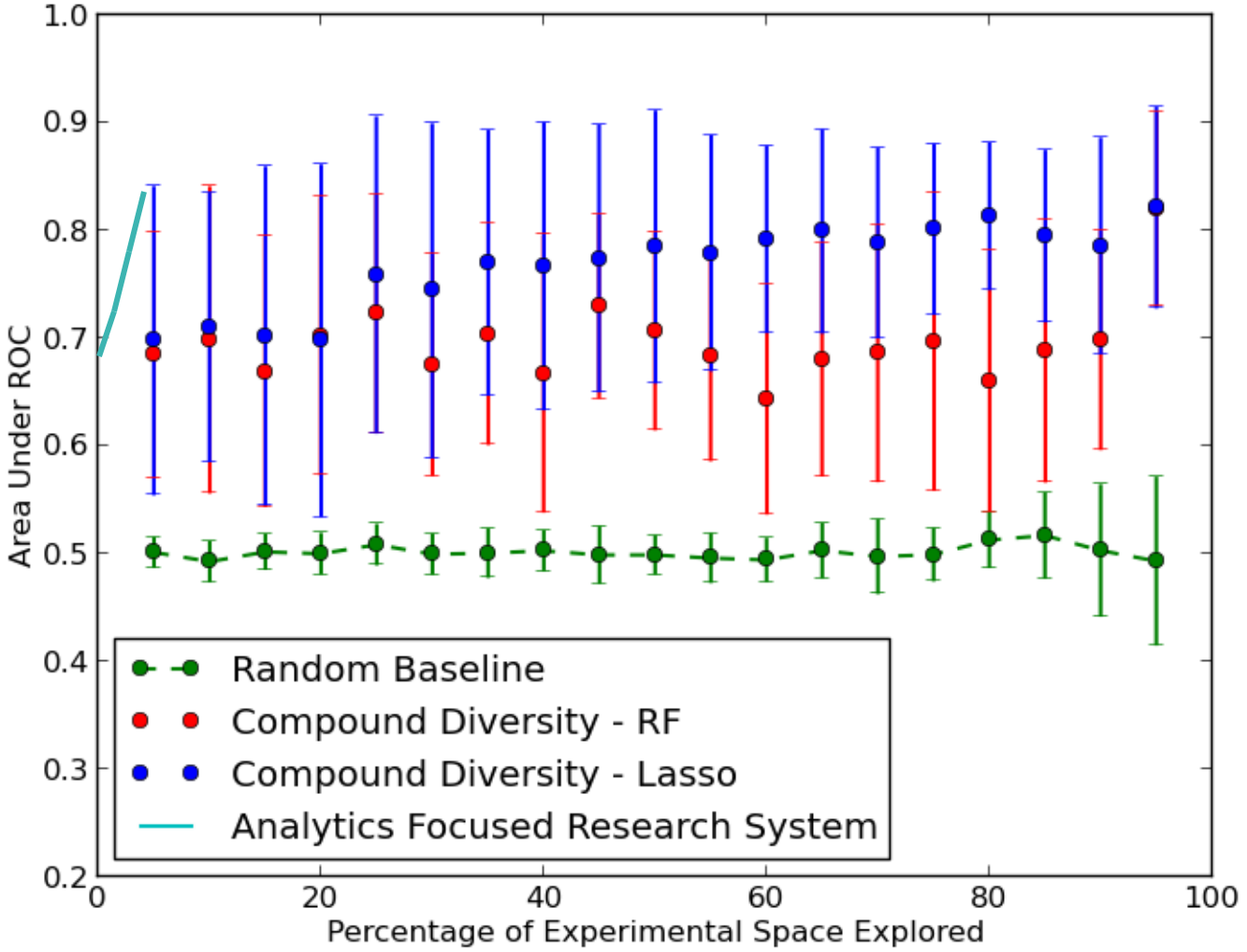

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Area Under ROC
0
20
40
60
80
100
Percentage of Experimental Space Explored
Random Baseline
Compound Diversity - RF
Compound Diversity - Lasso



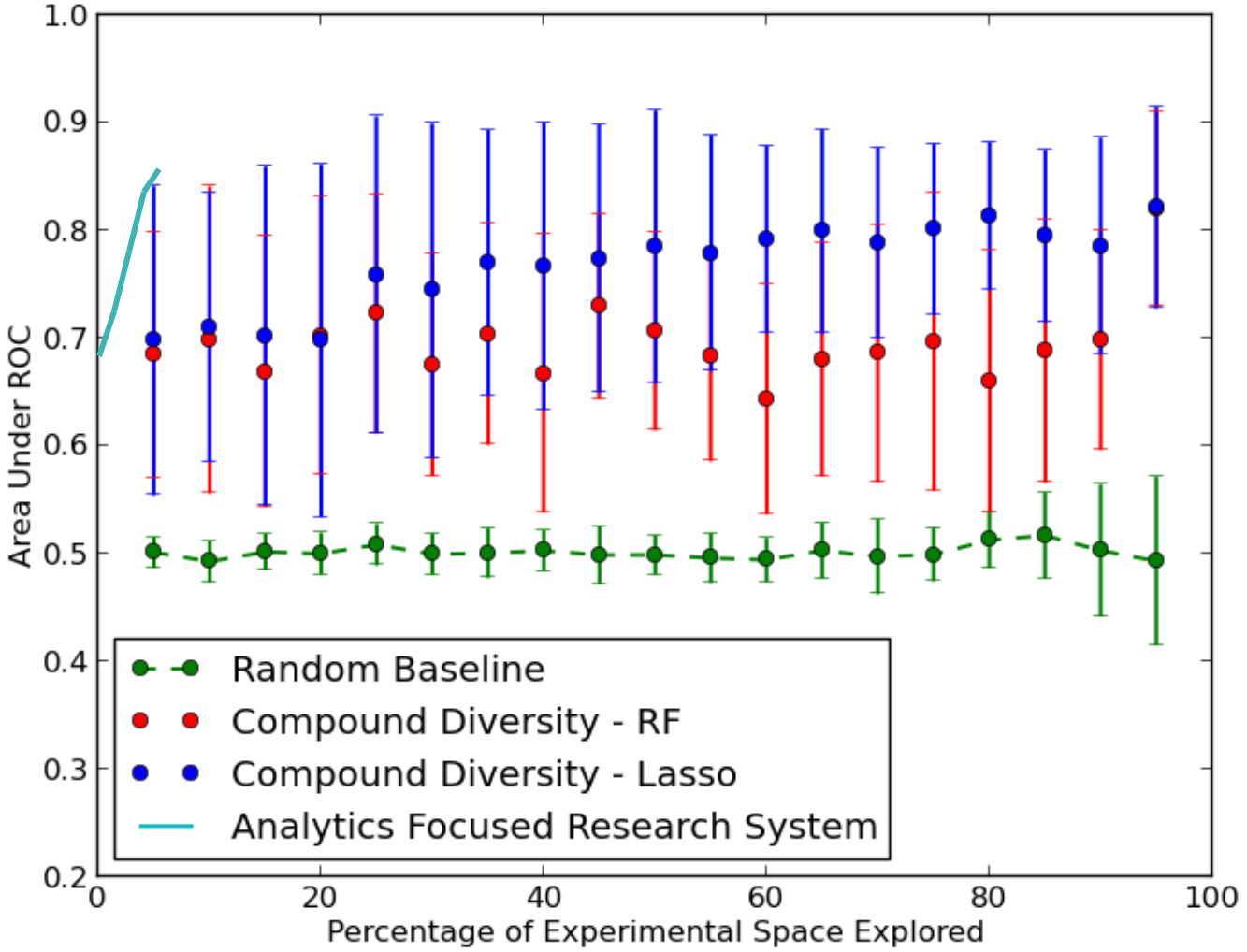
Active Learning Cycle Iteration: 1



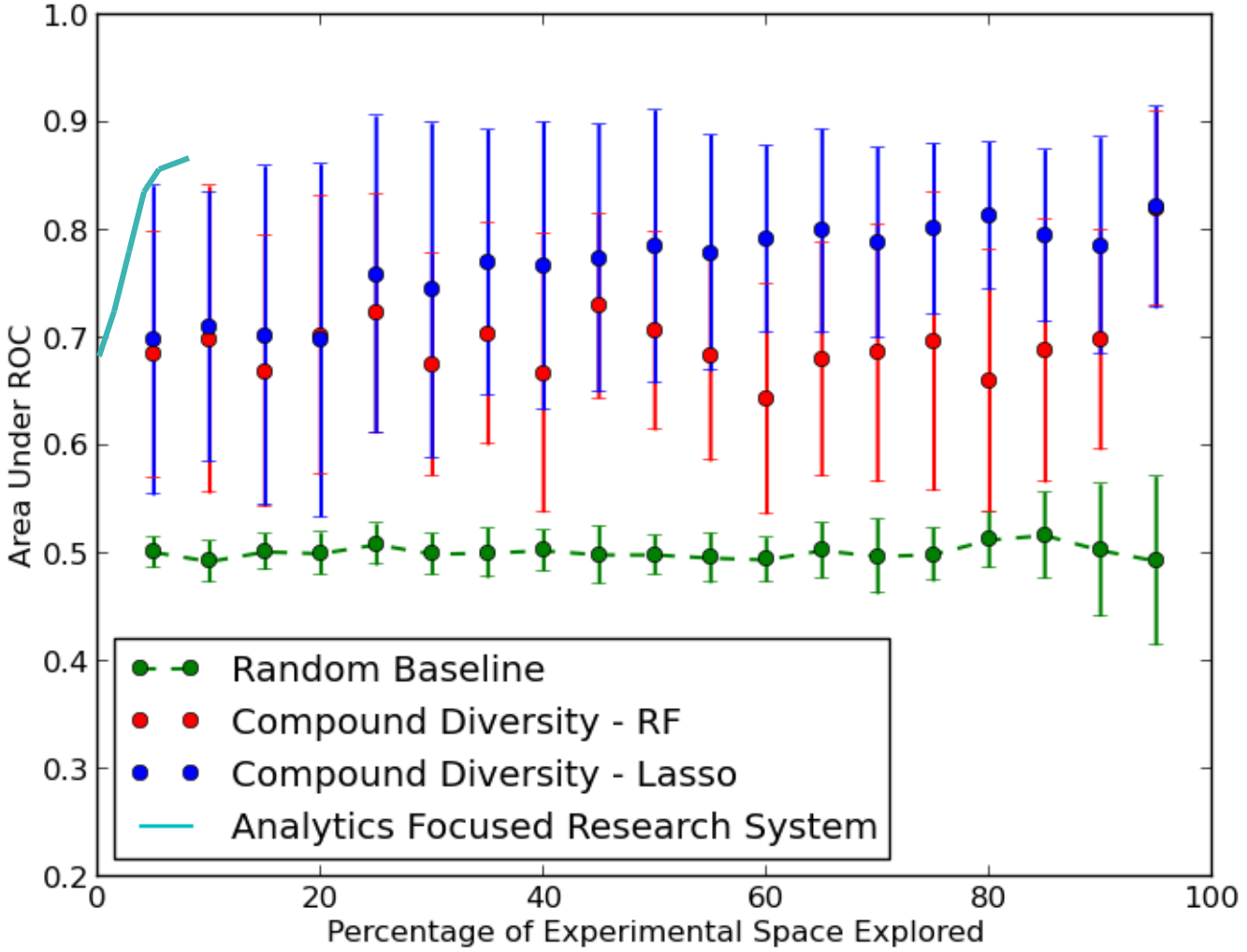
Active Learning Cycle Iteration: 3



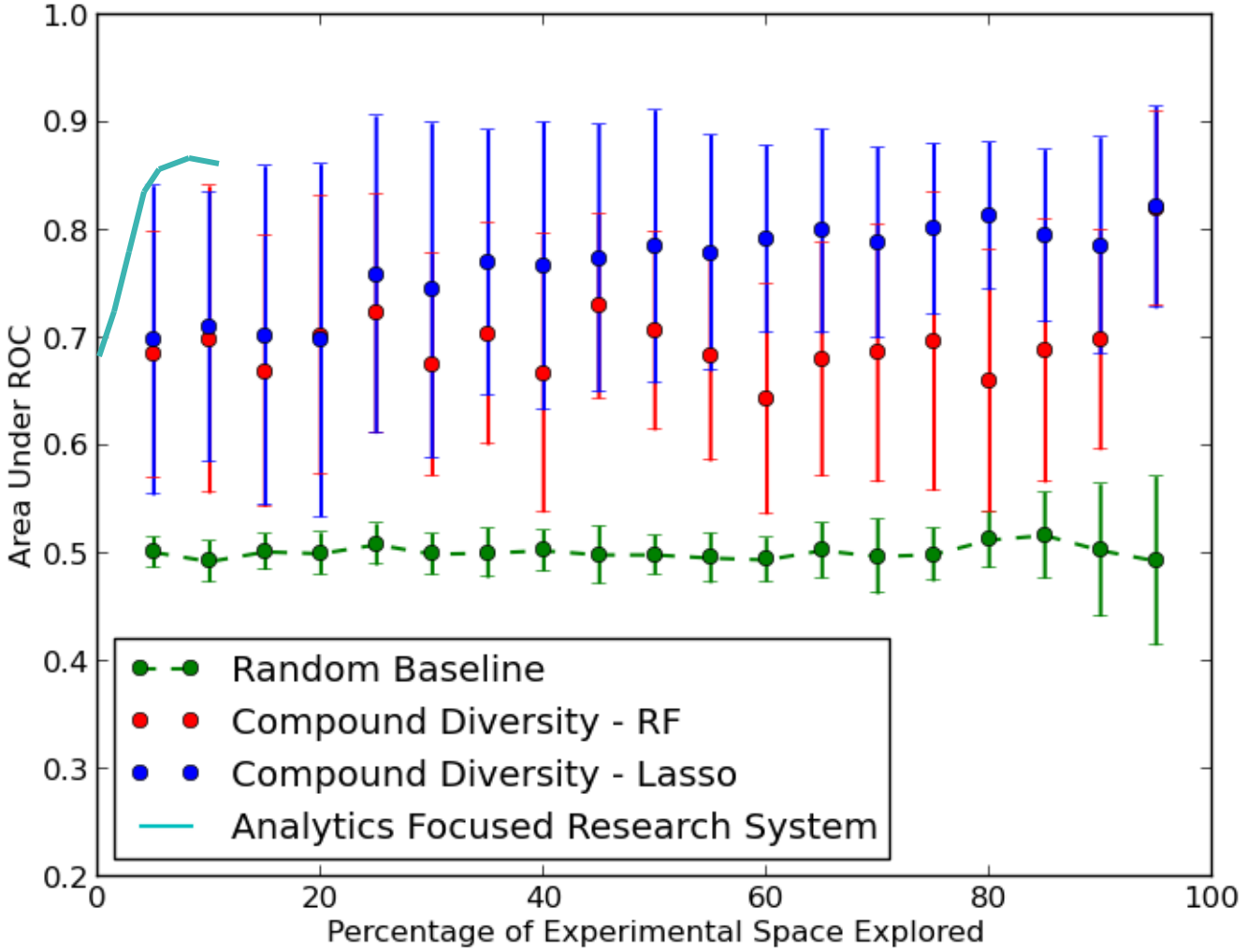
Active Learning Cycle Iteration: 4



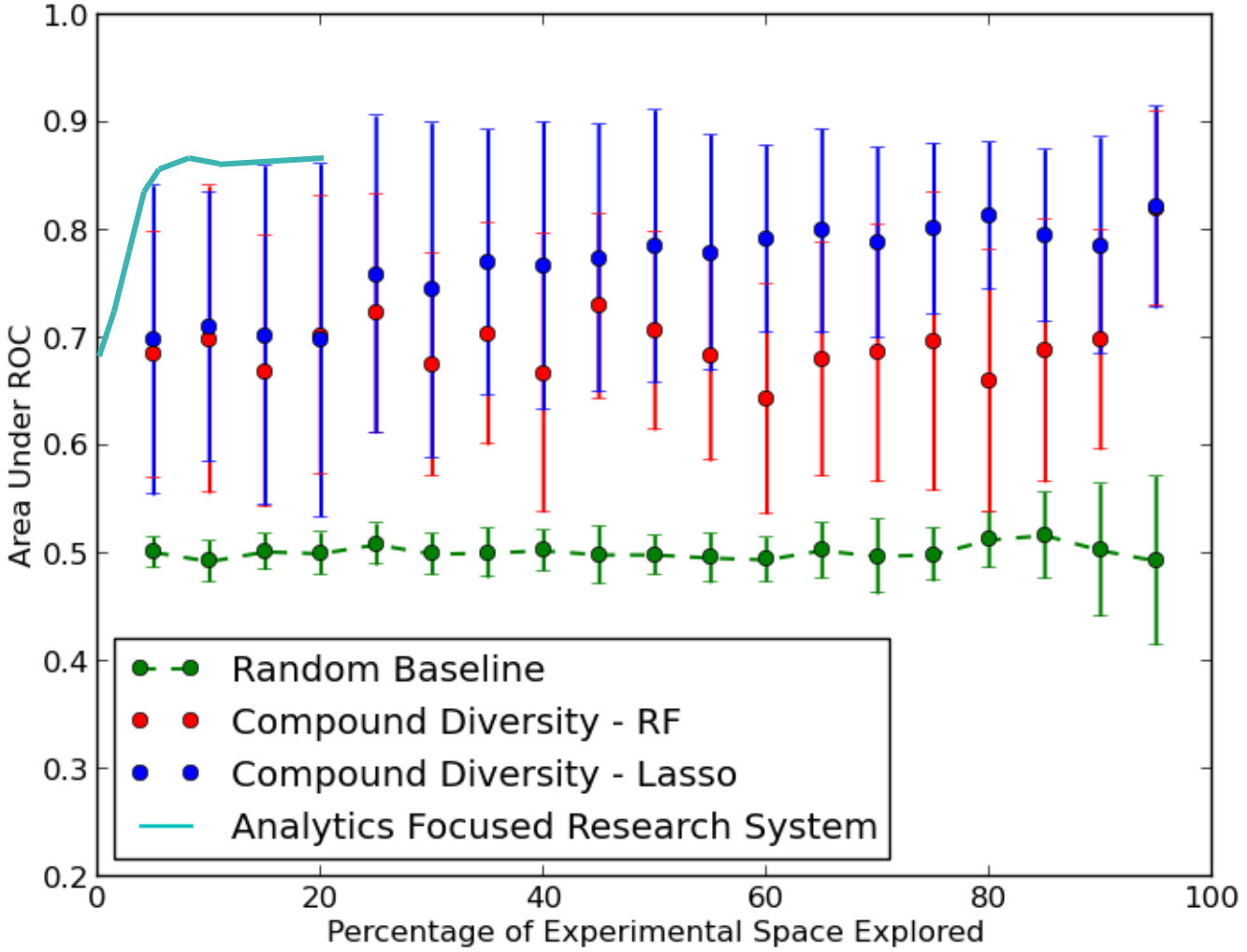
Active Learning Cycle Iteration: 6



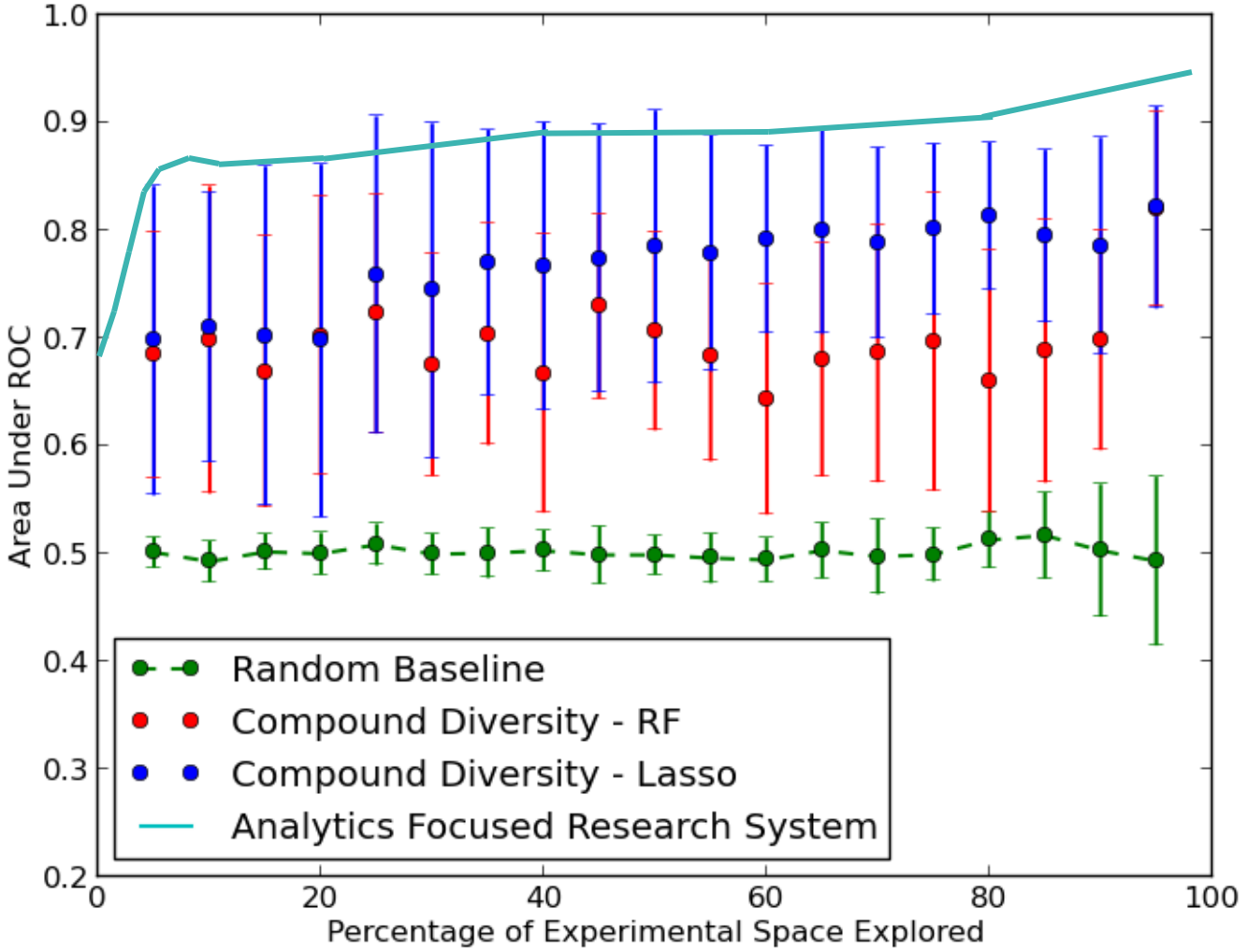
Active Learning Cycle Iteration: 8



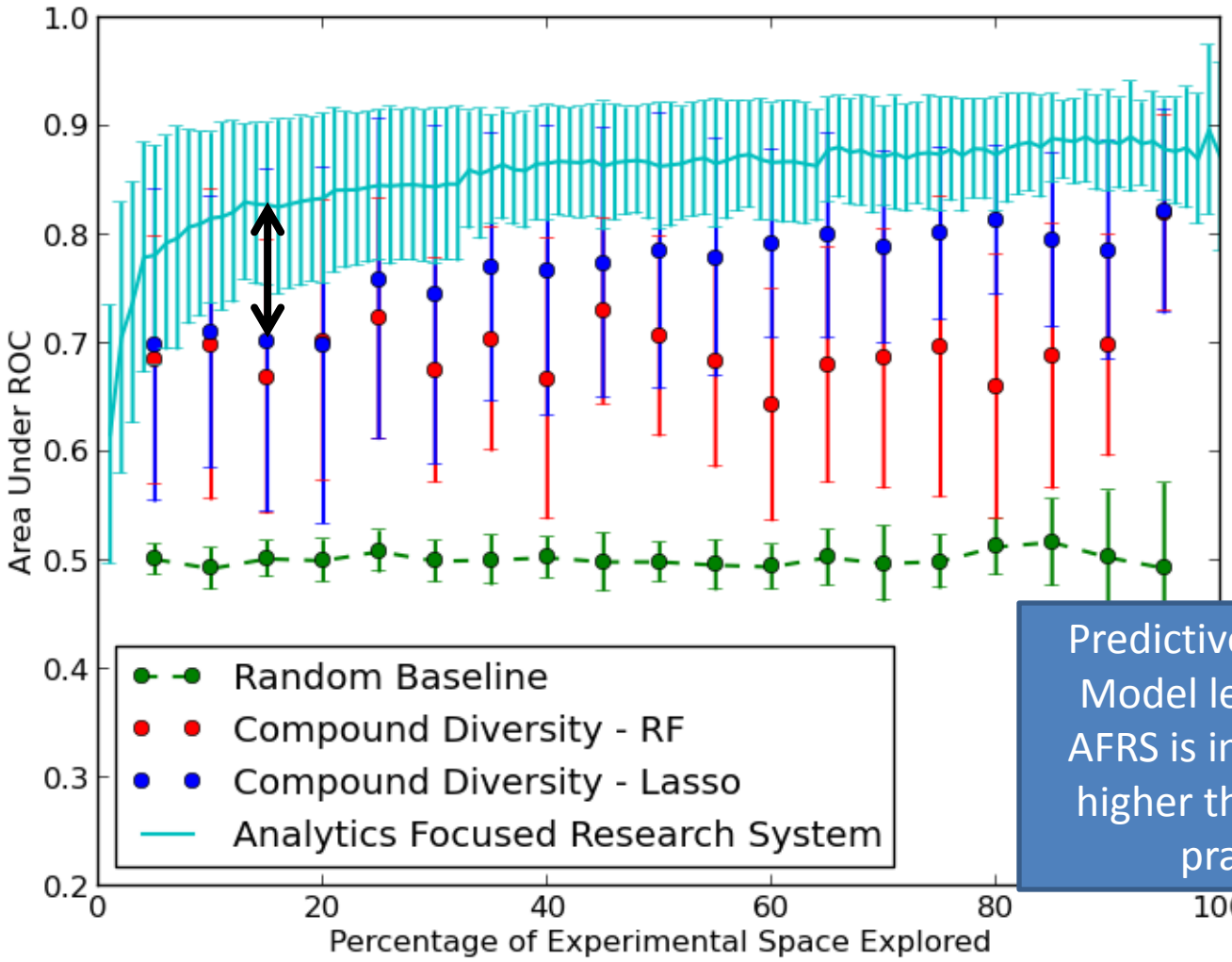
Active Learning Cycle Iteration: 11



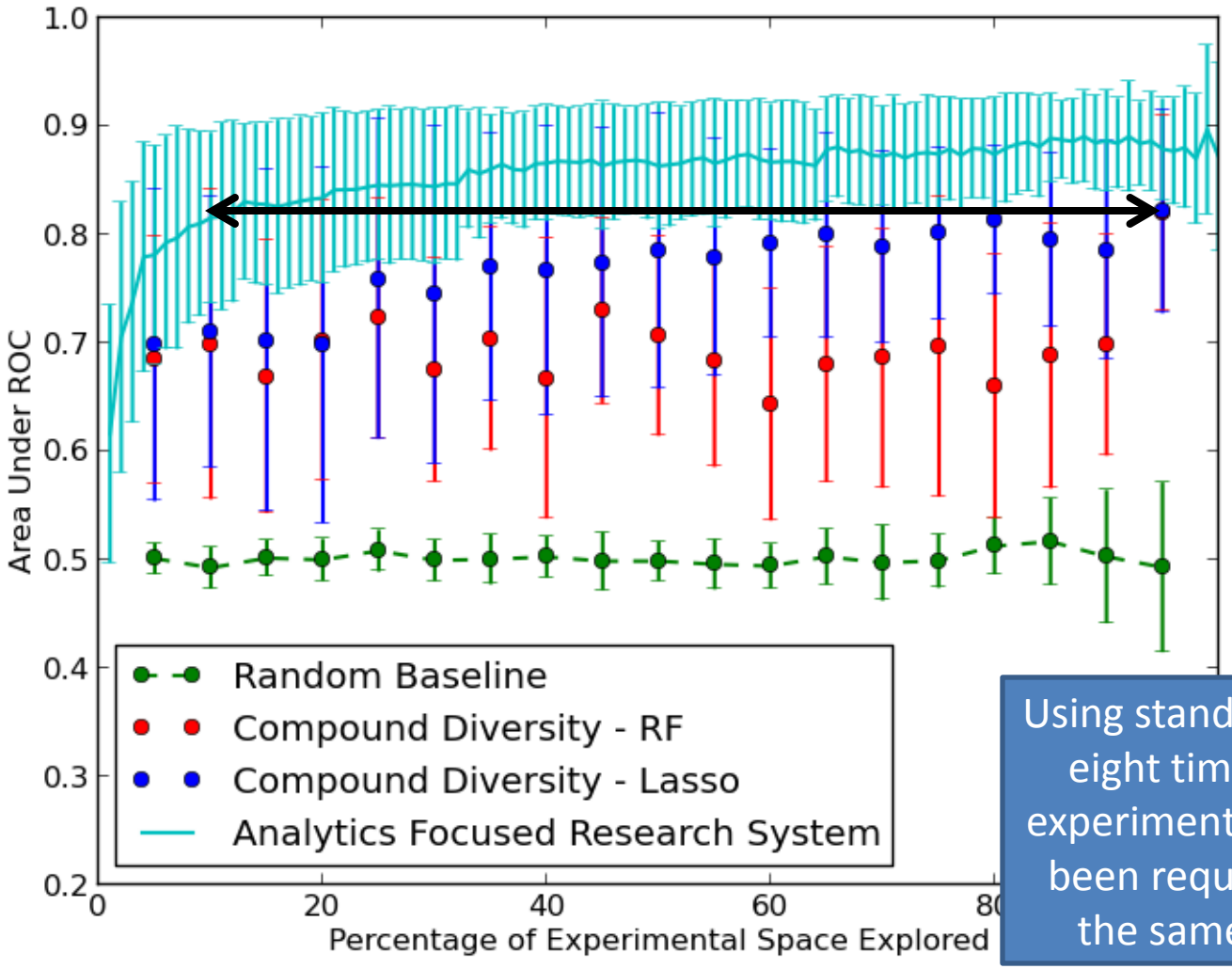
Active Learning Cycle Iteration: 20



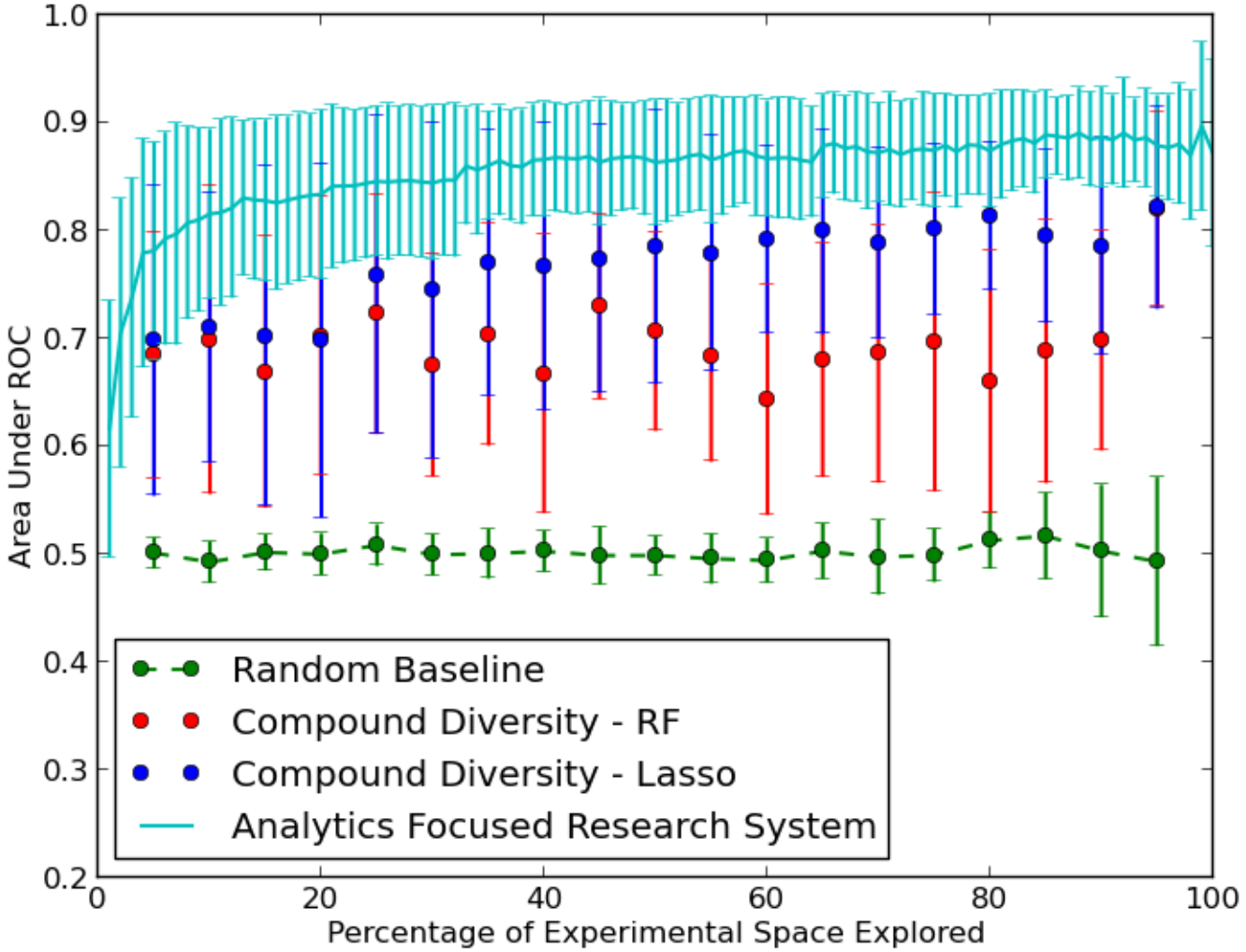
Active Learning Cycle Iteration: 99



Predictive accuracy of Model learned using AFRS is in expectation higher than standard practices.



Using standard methods, eight times as many experiments would have been required to have the same accuracy.


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Area Under ROC
0
20
40
60
80
100
Percentage of Experimental Space Explored
Random Baseline
Compound Diversity - RF
Compound Diversity - Lasso
Analytics Focused Research System

Example From a Single Simulation

- When predicting experimental results for Atrazine:

AFRS			Standard Approach		
Total Experiments Executed	Estimated Experimental Cost	Accuracy of Prediction for Atrazine (AUC)	Total Experiments Executed	Estimated Experimental Cost	Accuracy of Prediction for Atrazine (AUC)
32,000	$600,000	85%	260,000	$4,800,000	85%

ToxCast Phase 1 Estimated Total Cost: $6 million

Prospective Uses of AFRS Process

- For a small set of compounds:
  - Make predictions for toxicity
  - Prioritize toxicology assays to most effectively eliminate compounds (faster-to-fail)
  - Prioritize assays for improving predictive model

	emamectin benzoate	Aldicarb
Round 3	Novascreen - cytochrome P450, family 19, subfamily A, polypeptide 1 ELISA – chemokine (C-C motif) ligand 2	None
Round 7	Novascreen - ADME_hCYP4F12_Activator	Novascreen - cytochrome P450, family 19, subfamily A, polypeptide 1

Prospective Uses of AFRS Process

- For a new series of synthesized compounds:
  - Select a small number of experiments to effectively expand the predictions to new areas of chemical space as needed
  - Prioritize compounds based on predicted toxicity

Prospective Uses of AFRS Process

- For multiple assays:
  - Prioritize assays based on human toxicity predictivity
- For a large digital library of ***unsynthesized*** compounds
  - Select a small number of experiments from available ***synthesized*** compounds which give the most information about ***unsynthesized*** compounds
  - Prioritize compounds for synthesis
    - Based on predicted lack of toxicity
    - Based on potential information gain from testing synthesized compounds

Next studies with other data?

- Using your data, expand experimental space to include results from other areas of your company?
- Direct prediction of human toxicity with your data or alternative public source?
- Prospective study?

Quantitative Medicine

Transforming Drug Discovery

Scott R. Bodine
scott.bodine@qtmed.com

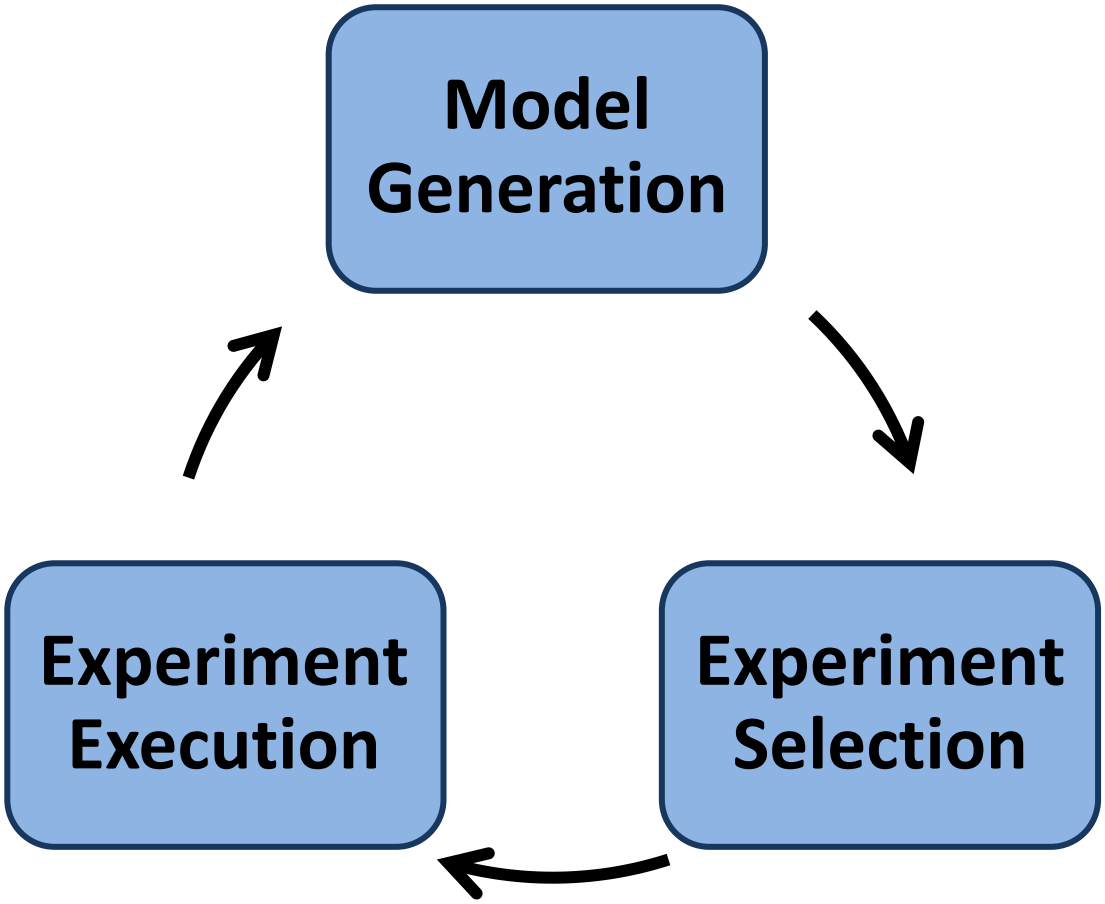
Robert F. Murphy PhD
robert.murphy@qtmed.com

Founders

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

Joshua D. Kangas PhD
josh.kangas@qtmed.com
Chief Science Officer

Active Learning



AFRS Directed Study Types

Retrospective

- Experiment Direction Simulated
- Show *potential* for improved **accuracy** and **efficiency**
- May discover new useful knowledge through AFRS analytics

Prospective

- Experiments Directed as Selected
- Yield *actual* improvements in accuracy and efficiency
- Likely to discover new useful knowledge through directed experimentation ***and*** AFRS analytics