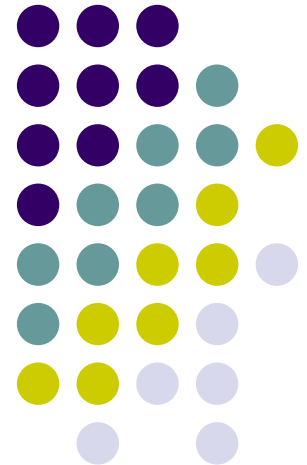


Genome-Wide Co-Expression Based Prediction of Differential Expressions

Yinglei Lai, Ph.D.

Department of Statistics & Biostatistics Center
The George Washington University

The 4th Annual Rocky Mountain Bioinformatics Meeting
Supported by **NIH/NIDDK DK-75004**





Data Sets

- Prostate cancer data set (Singh et al., 2002)
 - 50 normal and 52 cancerous
 - 6034 genes after thresholding and filtering
 - Many genes were identified
 - *Hepsin* and other novel genes
 - Several known disease related genes were not identified
 - *GSTP1*, *AMACR*
- Type 2 diabetes data set (Mootha et al., 2003)
 - 17 normal and 18 diabetic
 - 10983 genes after thresholding and filtering
 - No gene was identified



Motivation

- Student's t-test for differentiability measure
- Pearson correlation coefficient for correlation measure

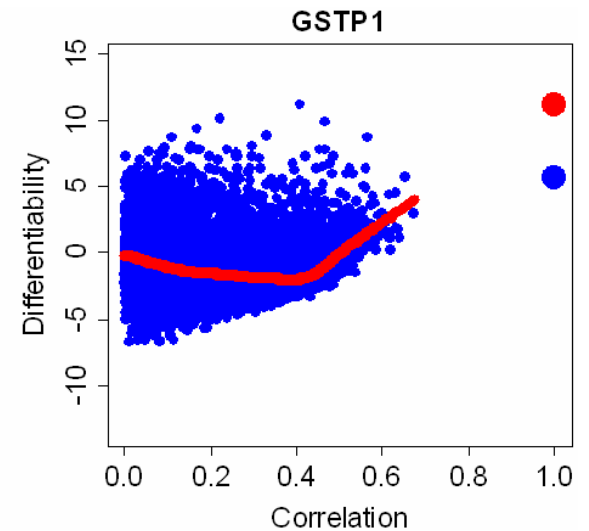
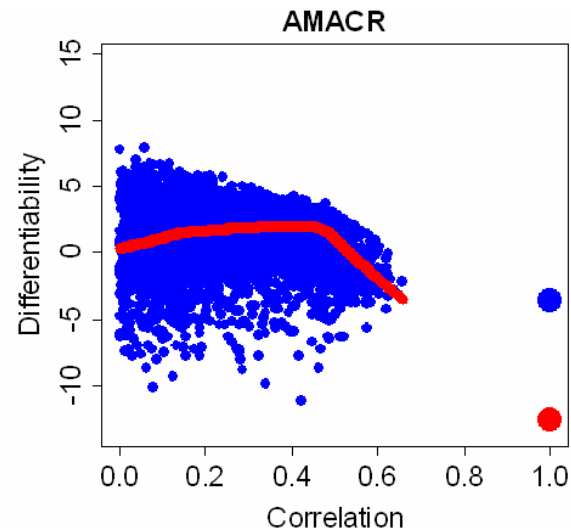
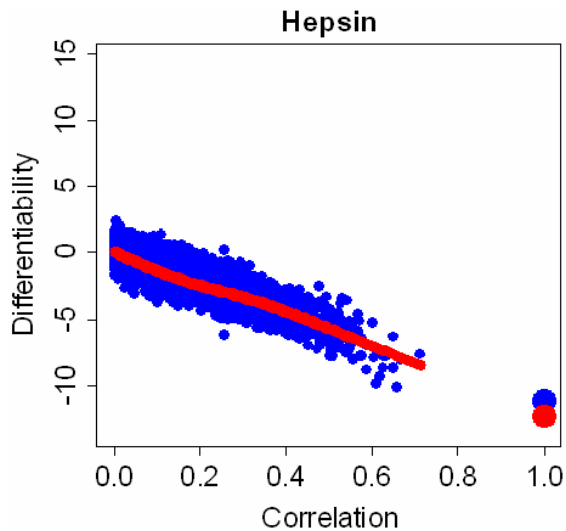
- Genes interact with each other during cellular processes
- The correlation and differentiability measures are related
- This relationship can be modeled by LOWESS
- This relationship can be used to predict the differentiability
- The significance can be evaluated by the permutation procedure

A mixture normal quantile based method for transforming data

LOWESS Prediction of Differentiability (Prostate Cancer Data Set)



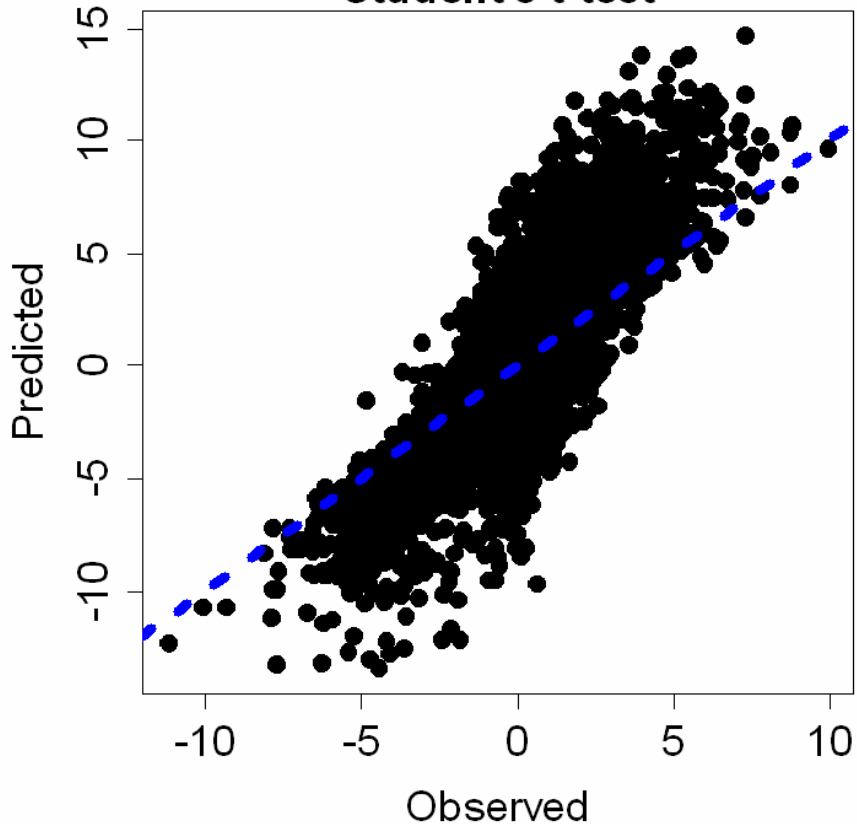
	<i>Hepsin</i>	<i>AMACR</i>	<i>GSTP1</i>
Observed	-11.2 (1)	-3.6 (672)	5.7 (106)
Predicted	-12.3 (15)	-12.6 (13)	11.2 (44)



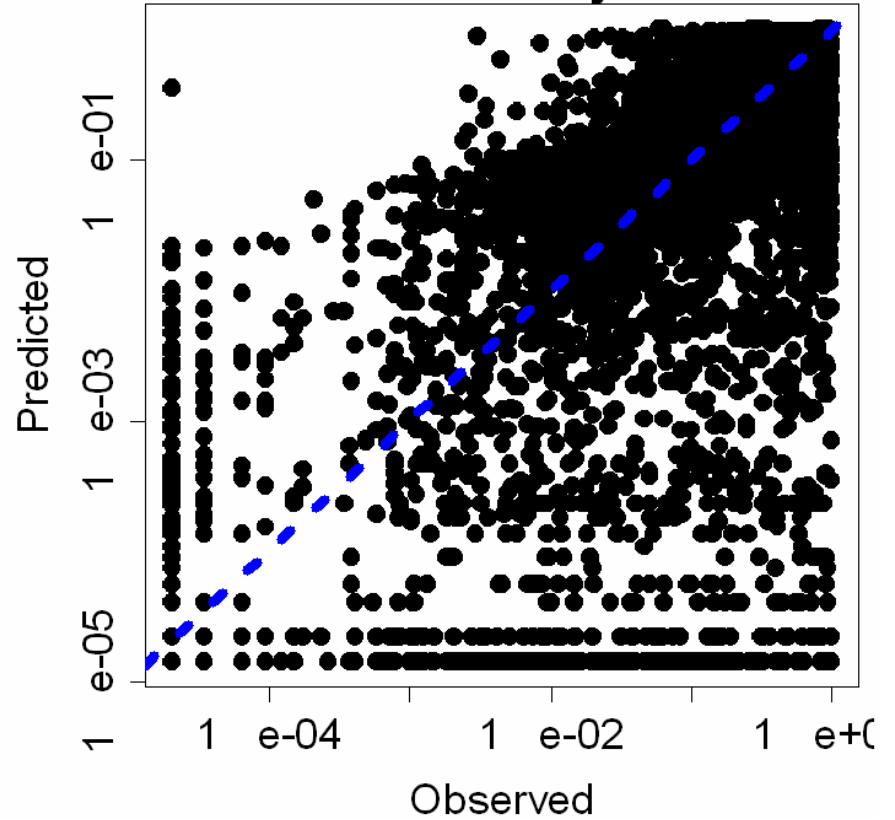
A Comparison (Prostate Cancer Data Set)



Student's t-test



False Discovery Rate



A Comparison (Type 2 Diabetes Data Set)

