

Modeling changes in gene expression in neurodegeneration in mice

Fourth Annual Primes MIT Conference

Lalita Devadas

Mentor: Angela Yen

May 18, 2014

Outline

1 Biological Background

2 Building Models

3 Results

4 Moving Forward

Outline

1 Biological Background

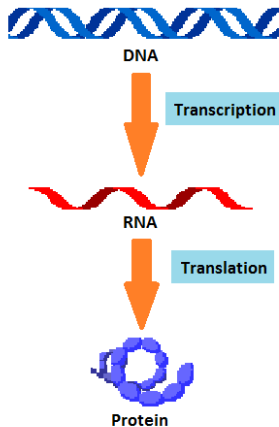
2 Building Models

3 Results

4 Moving Forward

Gene Expression

Central dogma of molecular biology



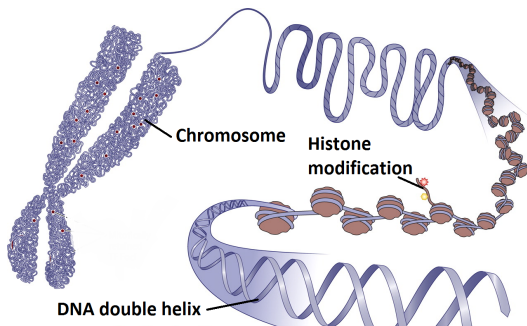
- Regulated by genetic (ACTG) and epigenetic factors

Epigenetics

Histone modifications

DNA Environment

Epigenetic factors are context that affect gene expression

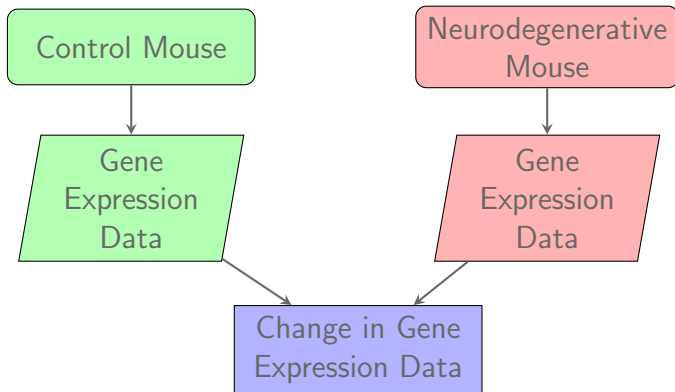


Histone Modifications

Chemical changes to histone protein core or protruding tail

Experimental Data

Neurodegeneration in mice



Outline

1 Biological Background

2 Building Models

3 Results

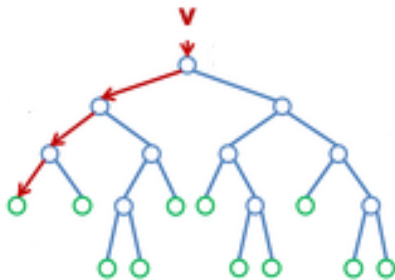
4 Moving Forward

Types of Models

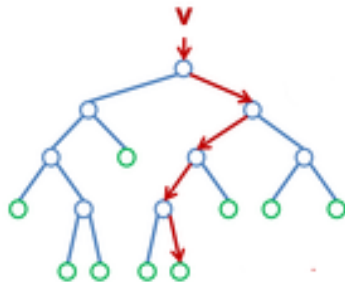
Random Forest

Random Forest model

Returns value based on set of values determined by a group of decision trees



Decision Tree 1



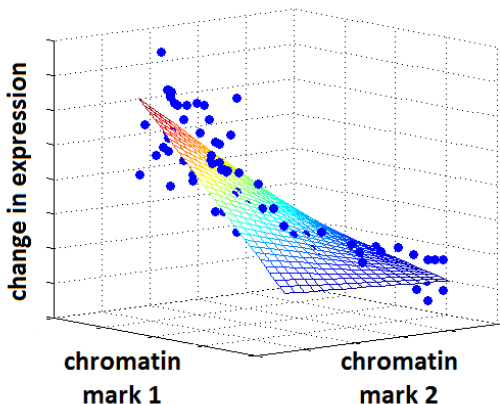
Decision Tree 2

Types of Models

Linear Model

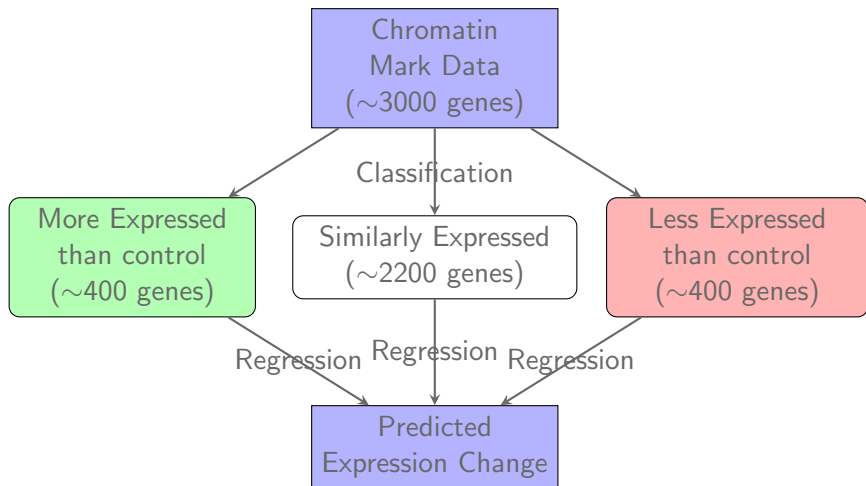
Linear model

Finds a linear correlation between predictors and response



Two-Step Model

Classification and Regression



Outline

1 Biological Background

2 Building Models

3 Results

4 Moving Forward

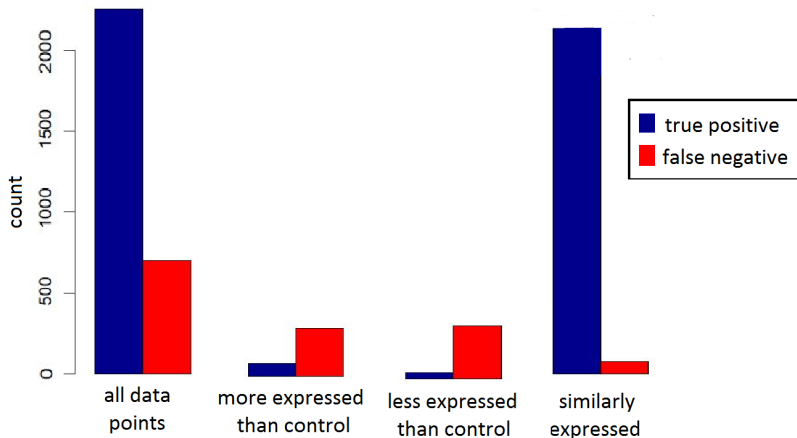
Final Model

- Train on half of data, test on other half
- Classification step: Random Forest model
- Regression step: Linear models

Results of Two-Step Model

Classification Graph

Accuracy of Classification



Results of Two-Step Model

Classification Values

Expression	Sensitivity	Specificity
More than control	.199	.043
Less than control	.063	.043
Same as control	.971	.128

Sensitivity

how good the model is at predicting if a data point belongs in a certain class

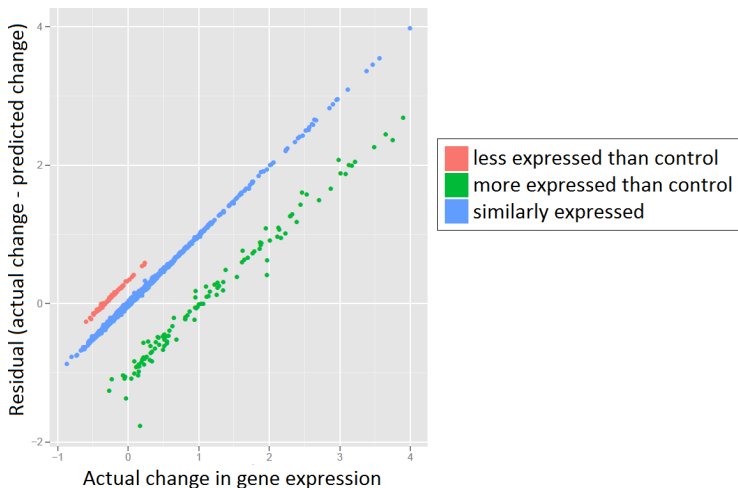
Specificity

how good the model is at predicting if a data point doesn't belong in a certain class

Results of Two-Step Model

Regression Graph

Accuracy of Regression: $r^2 \approx 0.18$



Outline

- 1 Biological Background
- 2 Building Models
- 3 Results
- 4 Moving Forward**

Next Steps

- Reprocess data (to improve our predictive power)
- Use different data (possibly Roadmap data)
- Create R package (for cross validation)

Acknowledgements

I would like to thank:

- My mentor, Angela Yen
- Prof. Manolis Kellis
- Andreas Pfenning
- Prof. Li-Huei Tsai and Elizabeth Gjoneska (Experimental collaborators)
- PRIMES program
- My family