Lab Meeting

Luke Hayden

26th September 2018

Background

Project

Examine the effect of regeneration on the molecular age profile of *Parhyale* limbs

Designing codeset
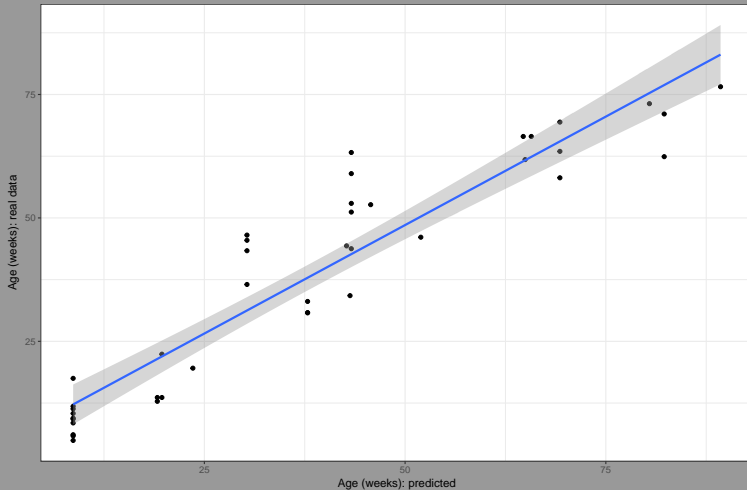
*Nanostring as method to quantify gene expression
*200 genes in codeset
-195 genes chosen on the basis of differential expression analysis
-5 control genes: do not vary in expression between conditions
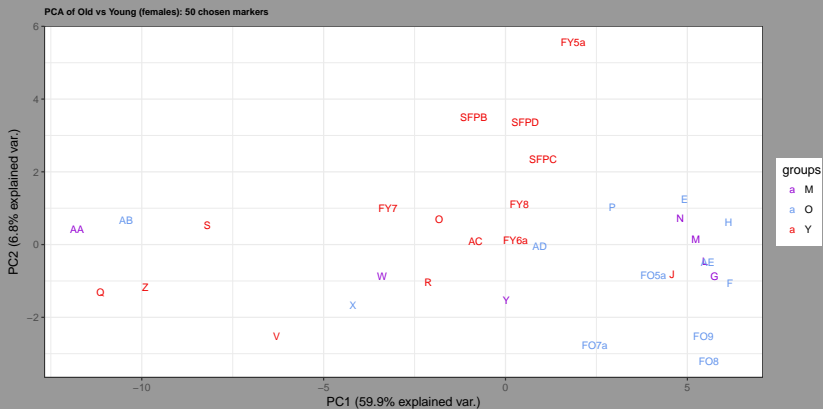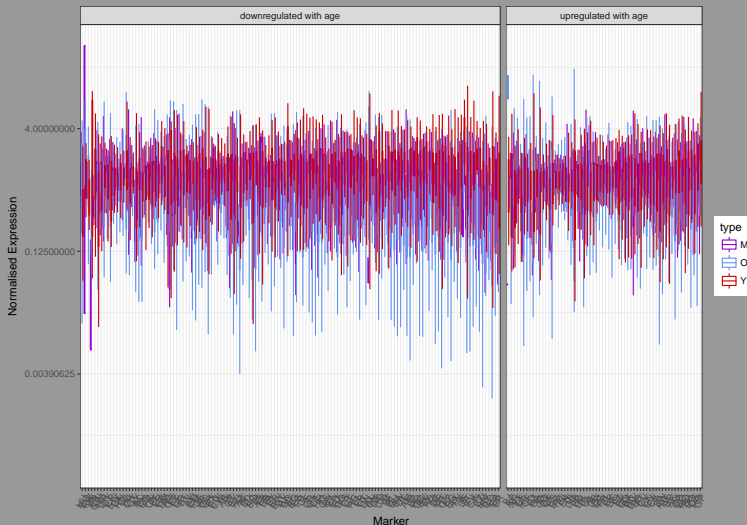
Age-length relationship

Age–length relationship in female Parhyale
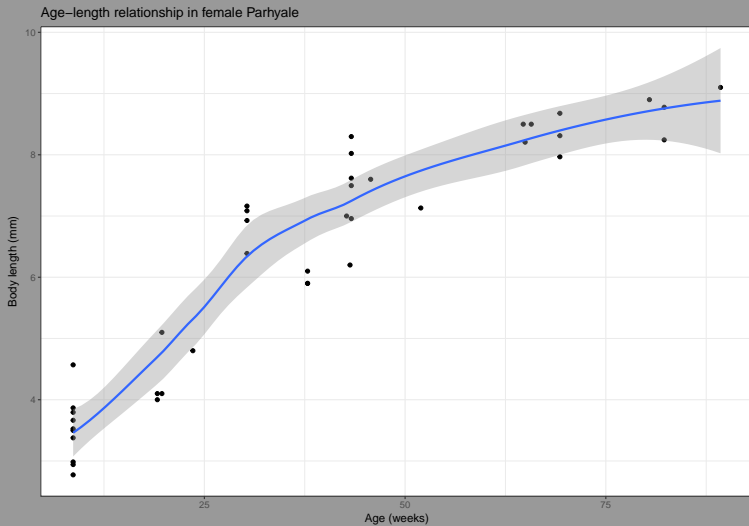
Building an age-length model

Model predicted vs real values

Young vs old separation: PCA old vs young

PCA of Old vs Young (females): 50 chosen markers

Young vs old separation: Old vs young by marker

Expression/length relationship
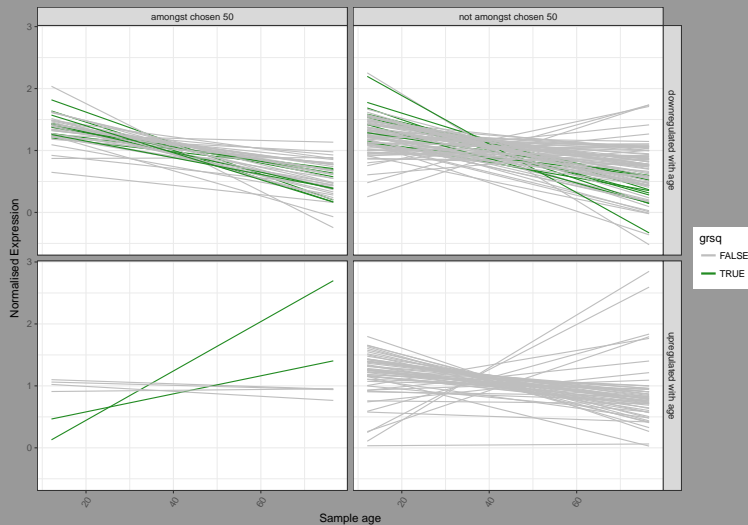
Multiple Regression approach

Simple linear regression:

Age = X(marker1) + c

We try to find values for x & c that come as close as possible to solving the equation for each set of values for *Age* and *marker1* we have.

Two predictors:

Age = X(marker1) + Y(marker2) + c

Many predictors

Age = X(marker1) + Y(marker2) + Z(marker3) + W(marker4) + ... + c

Where we have many different markers, we can find values of x,y,z,w, etc that solve this equation very well but don't provide predictive power

Random Forest approach

Decision tree
Classify or perform regression by asking binary questions of data: whether value of marker X is above or below key value Y, whther marker Z is above or below.....
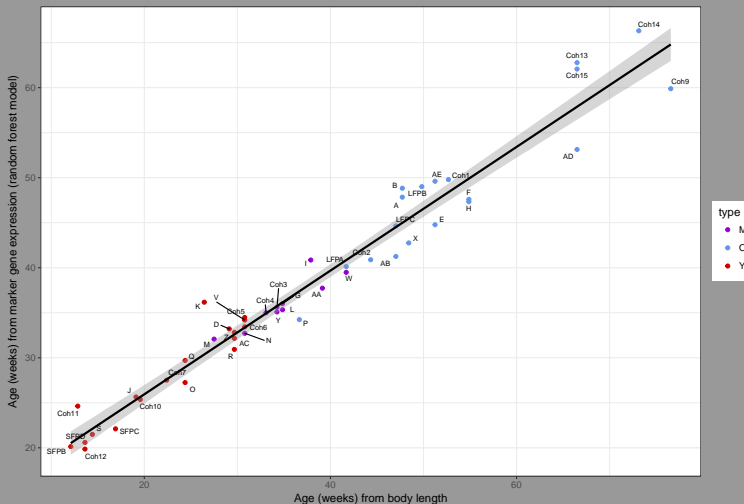
Random Forest
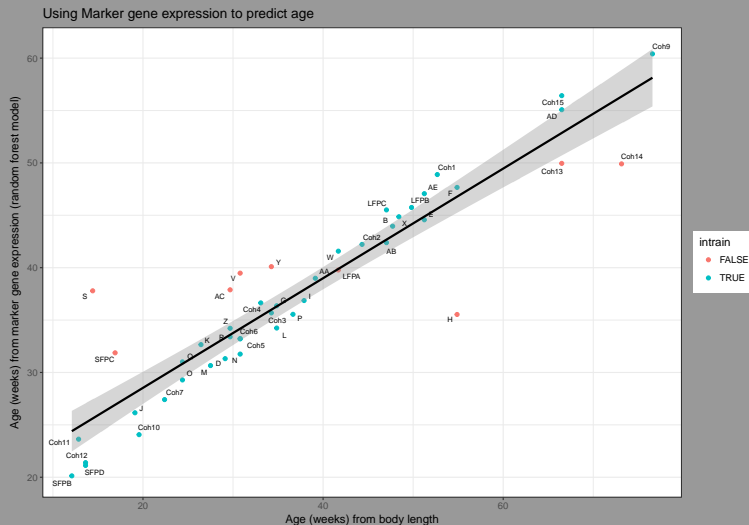Ensemble of decision trees, each using a random subset of the predictors
Resists overfitting

Examining this model

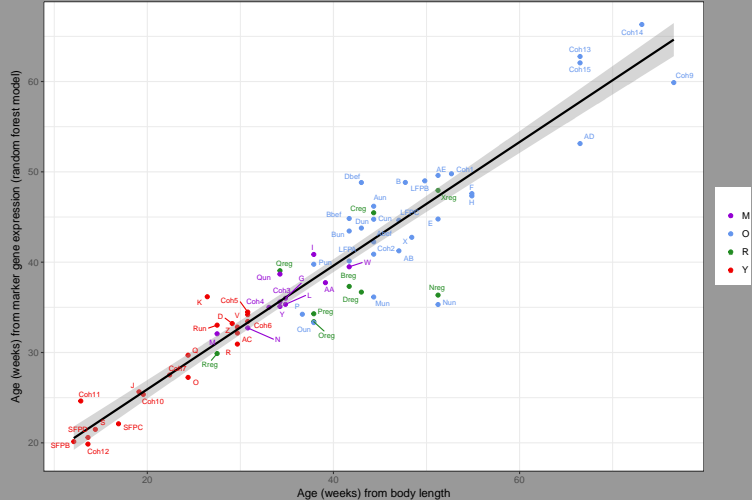Using Marker gene expression to predict age

Out-of-sample errors

Using Marker gene expression to predict age

Effect of regeneration

How does regeneration affect molecular age?

Testing this effect

Variably significant or not depending on details of model (choice of markers used, etc)

```
##
##  One Sample t-test
##
## data:  md$lmresid[md$reg == "R"]
## t = -1.4697, df = 8, p-value = 0.08992
## alternative hypothesis: true mean is less than 0
## 95 percent confidence interval:
##        -Inf 0.5763361
## sample estimates:
## mean of x
## -2.172771
```

New questions

How to optimise model?
Need to reduce out of sample error

Better way of ensuring sample quality?