

# Transcriptome Variation Under Dynamic Growth Conditions

or: How I learned to stop worrying and love RNAseq

Kevin Murray

Borevitz Lab

September 20, 2013

# Project Overview

- ▶ What I've spent my year doing:

# Project Overview

- ▶ What I've spent my year doing:
- ▶ Create & implement “dynamic growth conditions”

# Project Overview

- ▶ What I've spent my year doing:
- ▶ Create & implement "dynamic growth conditions"
- ▶ Study plant growth under these conditions

# Project Overview

- ▶ What I've spent my year doing:
- ▶ Create & implement "dynamic growth conditions"
- ▶ Study plant growth under these conditions
- ▶ Design & implement data analysis pipeline for RNAseq

# The Cazzonelli Button

*“Can’t there just be a ‘do my bioinformatics’ button?*

# The growth condition dilemma



- ▶ Plants grow in nature (mostly)

# The growth condition dilemma



- ▶ Plants grow in nature (mostly)



- ▶ Scientists work in labs (mostly)

# The growth condition dilemma



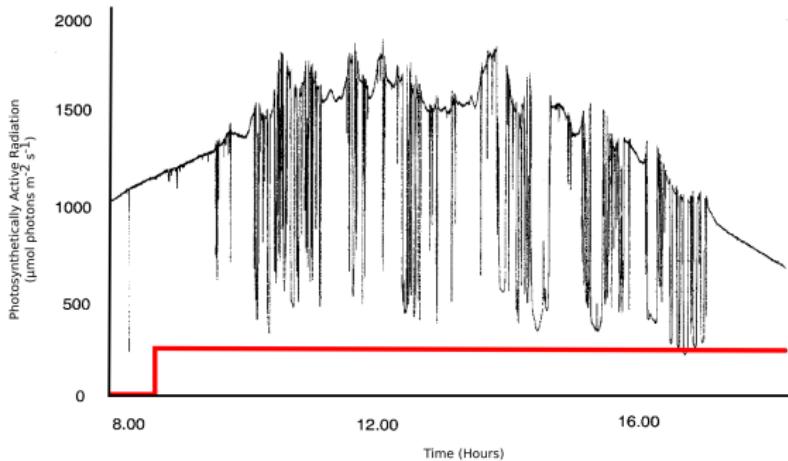
- ▶ Plants grow in nature (mostly)



- ▶ Scientists work in labs (mostly)

- ▶ Growth conditions have large effect on scientific results

# What are “Dynamic Growth Conditions”



- ▶ Model diurnal and circannual trends of climate
- ▶ Use this to govern lab growth conditions

# Light Quantity

- ▶ Plants need a “happy medium” of light
- ▶ Too little = suboptimal growth
- ▶ Too much = photooxidative damage

# Light Response & Transcriptomics

- ▶ Excess light invokes known transcriptional response

# Light Response & Transcriptomics

- ▶ Excess light invokes known transcriptional response
- ▶ Many genes induced in excess light constitutively expressed in field

# Light Response & Transcriptomics

- ▶ Excess light invokes known transcriptional response
- ▶ Many genes induced in excess light constitutively expressed in field
- ▶ Transcriptomics sensitive to subtly/fast changes

# Aims

1. Develop a pipeline of software to analyse RNAseq datasets

# Aims

1. Develop a pipeline of software to analyse RNAseq datasets
2. Determine the response of *Arabidopsis thaliana* to altered light intensity under dynamic growth conditions

# Aim 1

*Develop a pipeline of software to analyse RNAseq datasets*

# Aim 1

*Develop a pipeline of software to analyse RNAseq datasets*

- ▶ Define & overcome shortcomings in existing analysis pipelines
- ▶ Experimentally define limitations of RNAseq experimental designs

# What's a pipeline?

- ▶ A collection of software to turn raw data into results

# What's a pipeline?

- ▶ A collection of software to turn raw data into results



# What's a pipeline?



+



# What's a pipeline?



+



# Steps in an RNAseq Analysis

- ▶ Raw sequence data needs to be:
  - ▶ Filtered for quality
  - ▶ Aligned to the genome

# Steps in an RNAseq Analysis

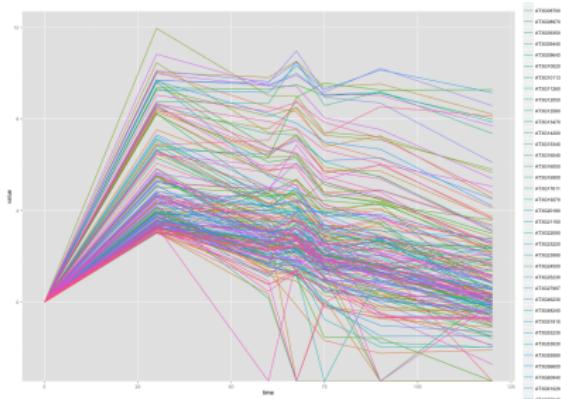
- ▶ Raw sequence data needs to be:
  - ▶ Filtered for quality
  - ▶ Aligned to the genome
- ▶ Once aligned, one needs to:
  - ▶ Quantify gene expression
  - ▶ Normalise counts
  - ▶ Perform statistical tests

## Existing “Best Practice” Pipeline

- ▶ Settings optimised for non plants
- ▶ Not optimised for large/dramatic changes
- ▶ Results in hypothesised weird artifacts

# Existing “Best Practice” Pipeline

- ▶ Settings optimised for non plants
- ▶ Not optimised for large/dramatic changes
- ▶ Results in hypothesised weird artifacts

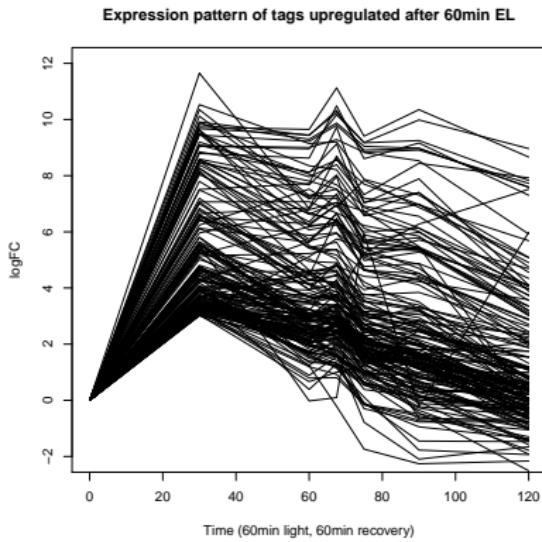


## “Improved” plot

- ▶ Using edgeR, state of the art statistics
- ▶ Hypothesis was artifacts would be removed

## “Improved” plot

- ▶ Using edgeR, state of the art statistics
- ▶ Hypothesis was artifacts would be removed

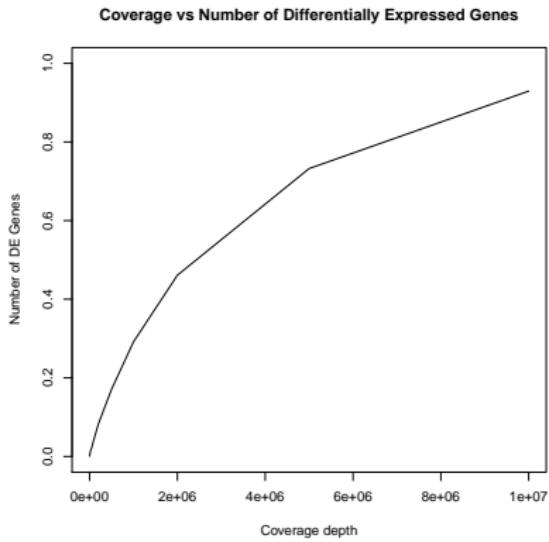


## Effect of Coverage

- ▶ Coverage depth a limiting factor in RNAseq
- ▶ Aimed to empirically determine “cutoff” depth for our study system

# Effect of Coverage

- ▶ Coverage depth a limiting factor in RNAseq
- ▶ Aimed to empirically determine “cutoff” depth for our study system



## Aim 2

*Determine the response of Arabidopsis thaliana to altered light intensity under dynamic growth conditions*

## Aim 2

*Determine the response of Arabidopsis thaliana to altered light intensity under dynamic growth conditions*

- ▶ Design & implement dynamic growth conditions

## Aim 2

*Determine the response of Arabidopsis thaliana to altered light intensity under dynamic growth conditions*

- ▶ Design & implement dynamic growth conditions
- ▶ Generate phenomic and transcriptomic QTL mapping datasets from plants grown under dynamic light conditions

## Aim 2

*Determine the response of Arabidopsis thaliana to altered light intensity under dynamic growth conditions*

- ▶ Design & implement dynamic growth conditions
- ▶ Generate phenomic and transcriptomic QTL mapping datasets from plants grown under dynamic light conditions
- ▶ Determine effect of light intensity on transcriptome under dynamic light conditions

# Introducing the SpectralPhenoClimatron

- ▶ Simulate regional climates
- ▶ Use model simulation to drive actual growth cabinet conditions
- ▶ Wrote python module spcControl



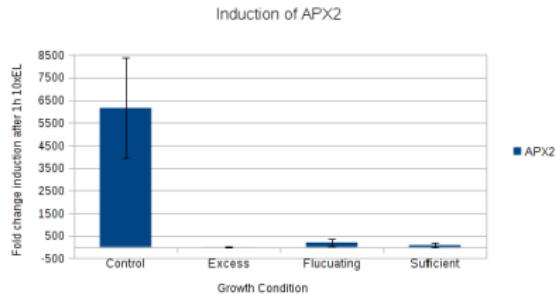
# Investigating altered light intensity

- ▶ Within a simulated climate, modify light intensity
- ▶ Create 3 new conditions:
  - ▶ Sufficient light
  - ▶ Fluctuating light
  - ▶ Excess light

plotofsolarcalcconditions.png

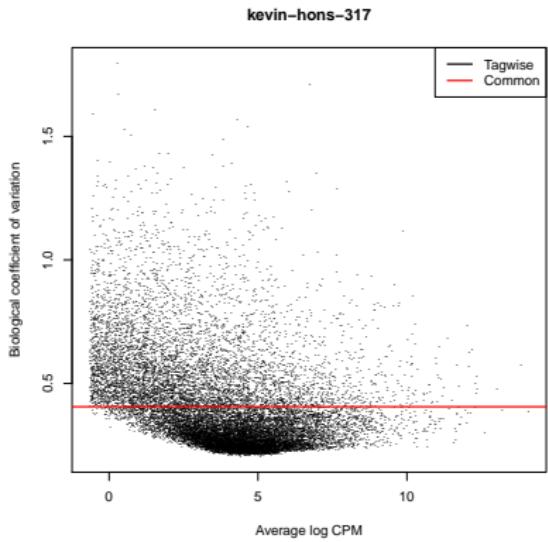
# qPCR analysis tells a small story

- ▶ Examine expression of known excess light responsive genes
- ▶ Show reduced induction and increased steady state expression



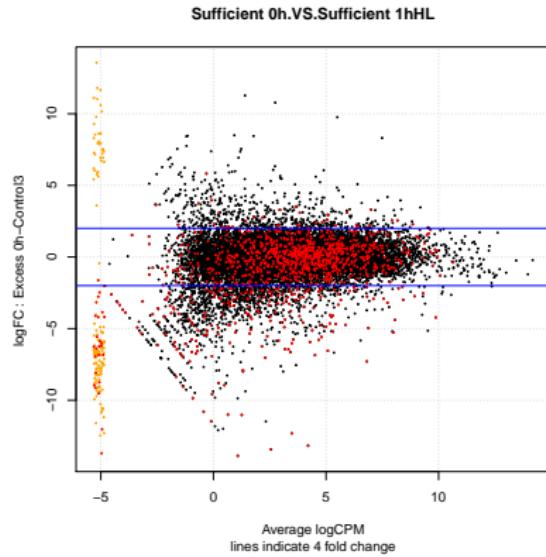
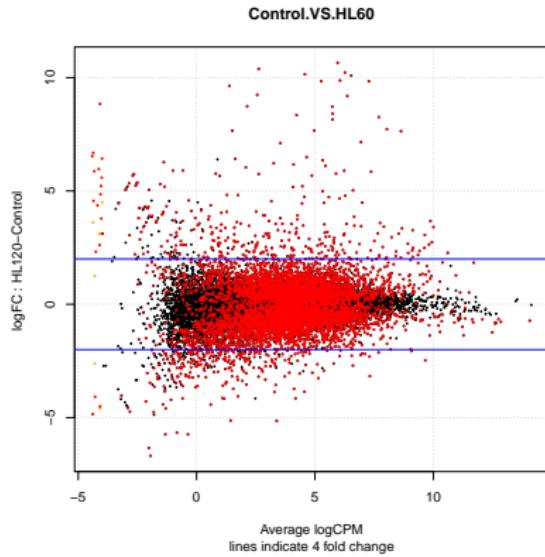
# RNAseq shows the whole picture

- ▶ Overall, high variation amongst samples



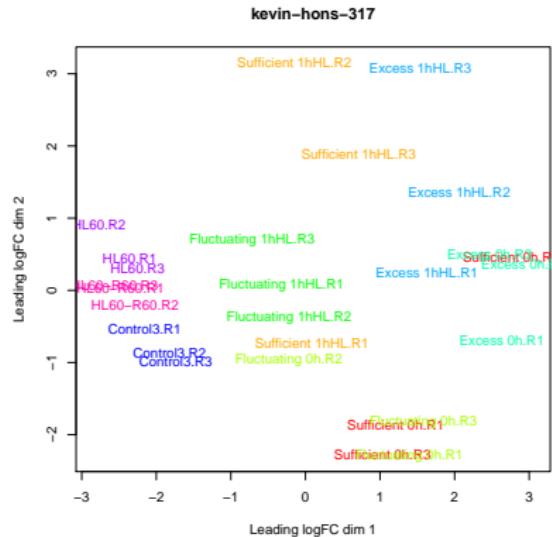
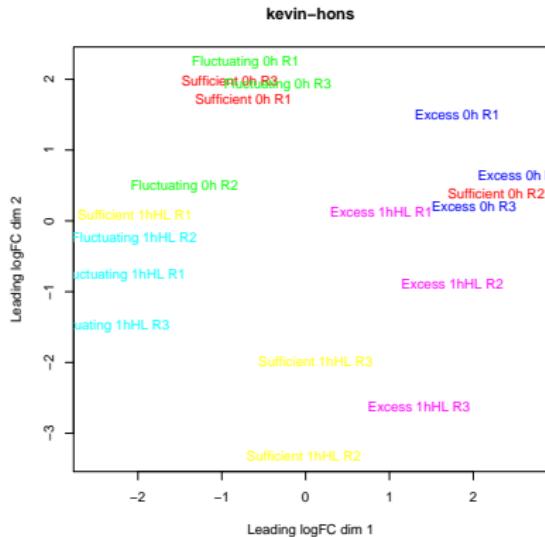
# RNAseq shows the whole picture

- Less differential expression than expected



# RNAseq shows the whole picture

- ▶ Amongst a noisy response, a pattern emerges



# Conclusions

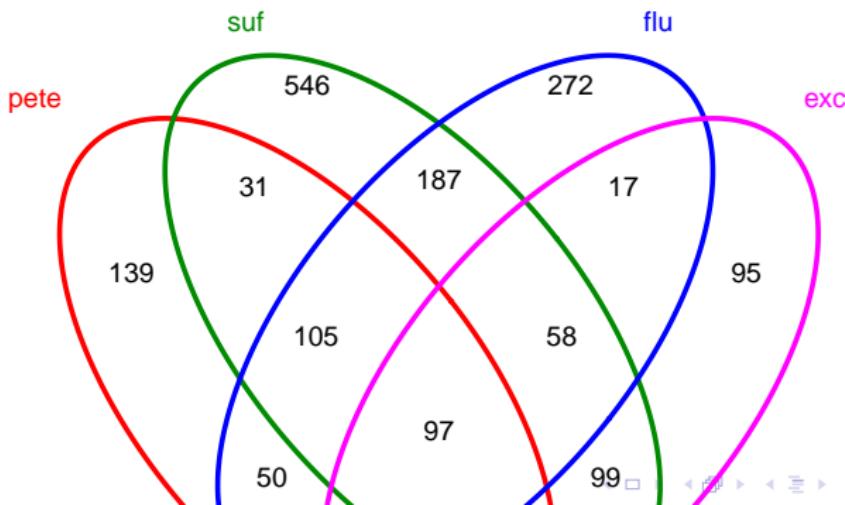
- ▶ Dyanmic growth conditions give intersting biology
- ▶ Patterns of differential expression seen, more replicates needed
- ▶ RNAseq is here, and easier than you might think



# Venn Diagrams

- ▶ Hard to see, but there is most overlap between Sufficient and Control, and Sufficient and Fluctuating

## Similarity in DTags



# Growth of plants

- ▶ A QTL mapping set, parental lines and genetic controls
- ▶ 3 replicates
- ▶ **Images forthcoming**