# Transcriptome Variation Under Dynamic Growth Conditions

Kevin Murray

Borevitz and Pogson Labs

September 24, 2013

# Big Picutre Question

- A set of broad hypotheses:
  - In nature, plants acclimate to repeated stresses
  - After acclimation, plants will be more tolerant to stresses
  - We can recreate this acclimation in lab using dynamic growth conditions

# Terminology

- **Transcriptomics:** study of global gene expression
- **RNAseq:** transcriptome quantification by sequencing
- **Pipeline:** series of software which turns data into results
- **QTL Mapping:** technique to associate variation in genotype to phenotype variation

# Investigating altered light intensity

- Hypothesised "hardening" of plants to harsher conditions
  - Increased steady state expression of stress genes
  - Decreased induction of stress genes after stress
- Hypothesised a relative order of "hardening"
  1. Fluctuating light
  2. Excess light
  3. Sufficient light
  4. Standard growth conditions

# Aims

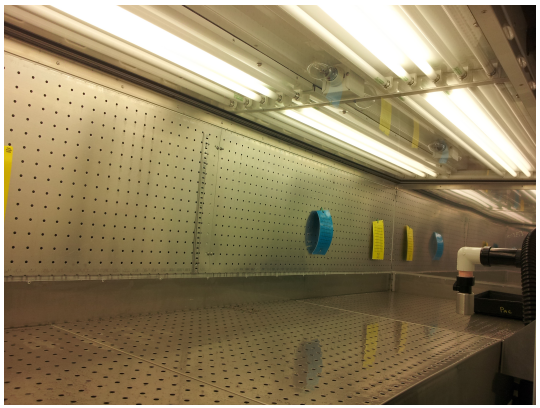
1. Design & implement dynamic growth conditions
2. Develop a pipeline of software to analyse RNAseq datasets
3. Generate phenomic and transcriptomic QTL mapping datasets from plants grown under dynamic light conditions
4. Determine effect of light intensity on transcriptome under dynamic light conditions

# The growth condition dilemma

- Plants grow in nature (mostly)

# The growth condition dilemma

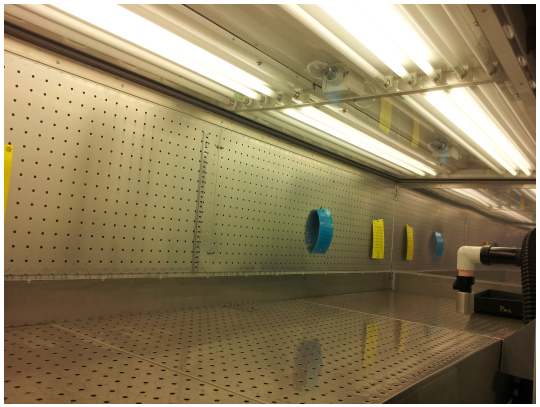
- Plants grow in nature (mostly)



- Scientists work in labs (mostly)

# The growth condition dilemma

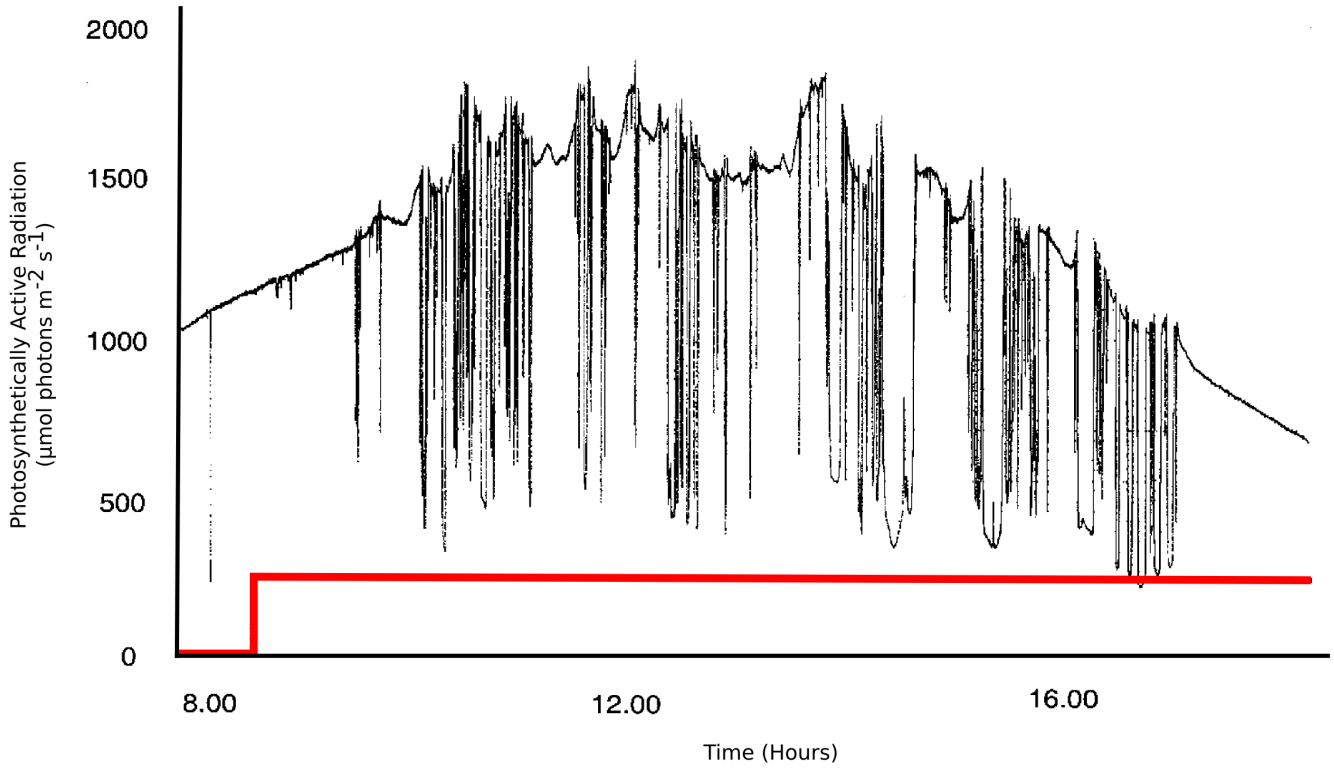
- Plants grow in nature (mostly)



- Scientists work in labs (mostly)

- Aim to “merge” elements of these two scenarios

# Introducing the SpectralPhenoClimatron




- Simulate regional climates
- Model diurnal and circannual trends of climate

(Külheim, Ågren, and Jansson 2002)

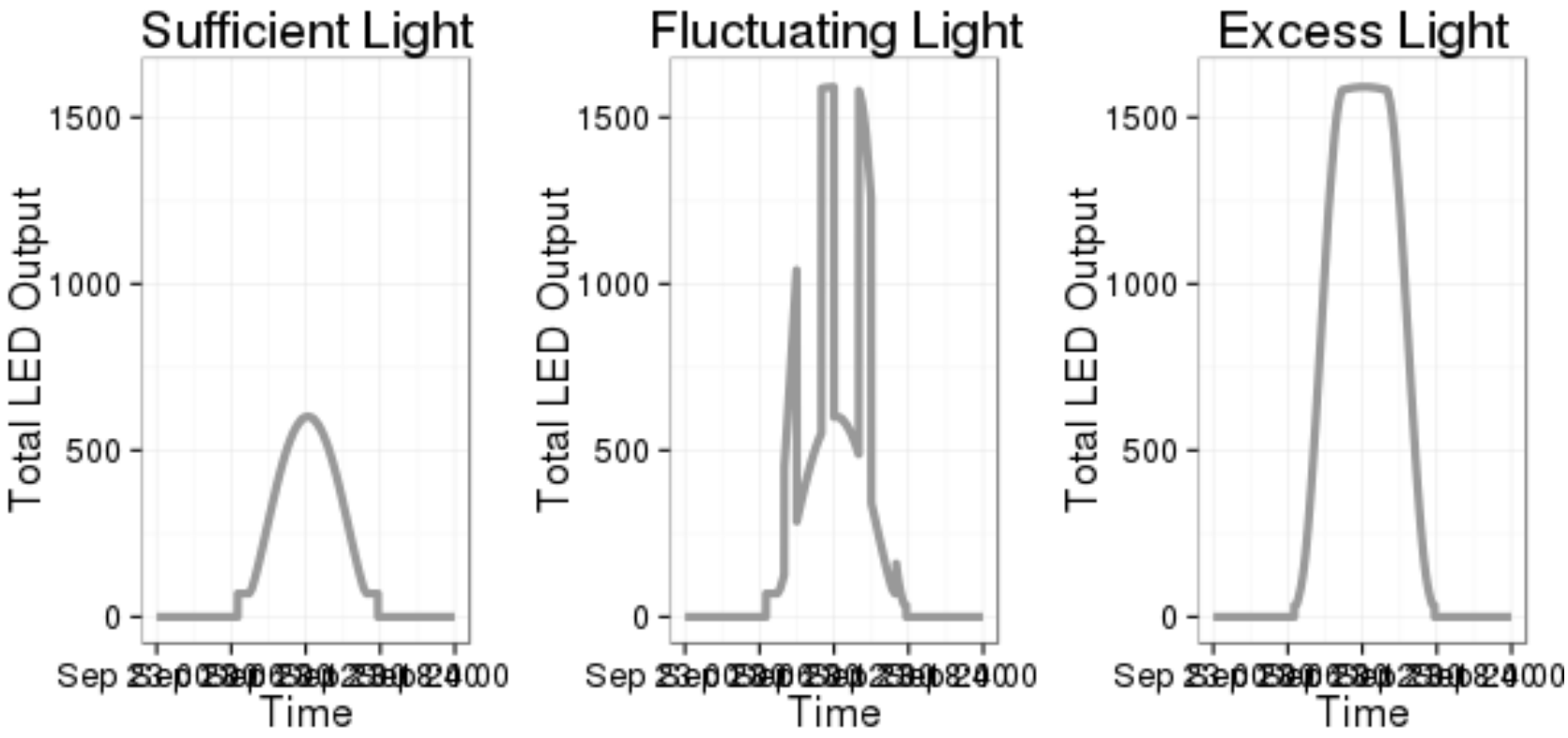
# Controlling the SpectralPhenoClimatron

- Use model simulation to drive actual growth cabinet conditions
- Need software "glue" to stick bits together
- Wrote `spcControl`
  - 750 lines
  - 134 minor, 16 major versions

Movie:

# Investigating altered light intensity

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



# Investigating altered light intensity

- Hypothesised "hardening" of plants to harsher conditions
  - Increased steady state expression of stress genes
  - Decreased induction of stress genes after stress
- Hypothesised a relative order of "hardening"
  1. Fluctuating light
  2. Excess light
  3. Sufficient light
  4. Standard growth conditions

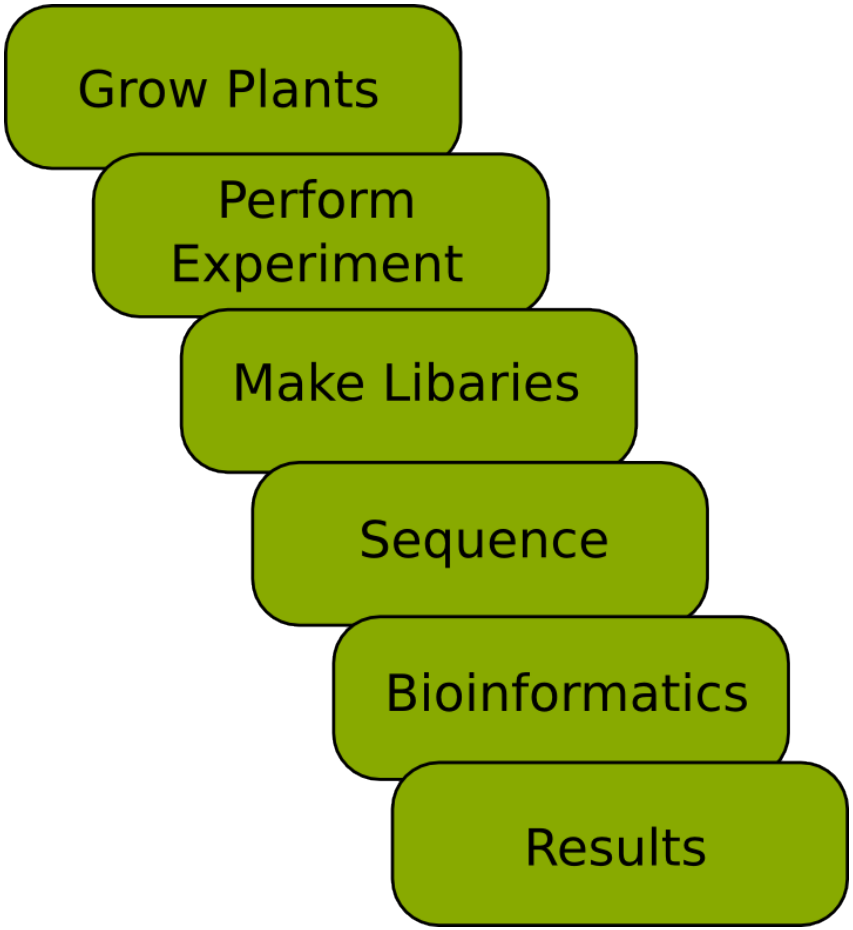
# Aims

- Design & implement dynamic growth conditions
- Develop a pipeline of software to analyse RNAseq datasets
- 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

# What's a pipeline?

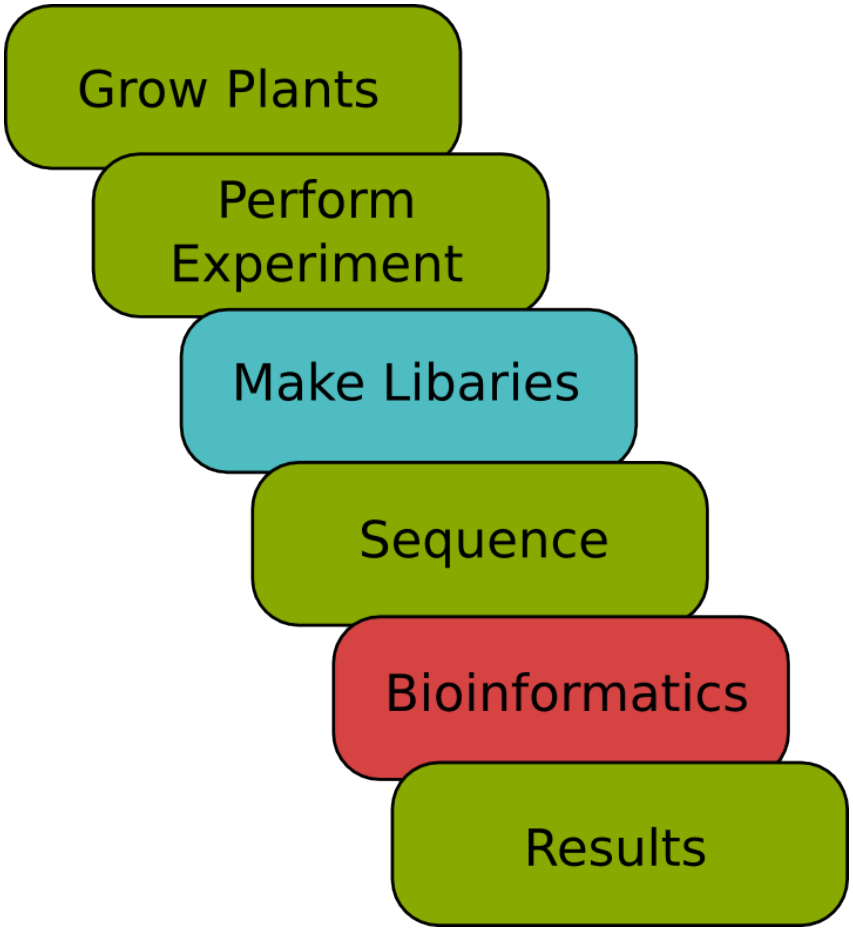
- A collection of software to turn raw data into results

# How does RNAseq work?




- Assay **ALL** expression in your tissue
- Unbiased, as quantitative as qPCR
- Becoming cheaper and easier

# How does RNAseq work?

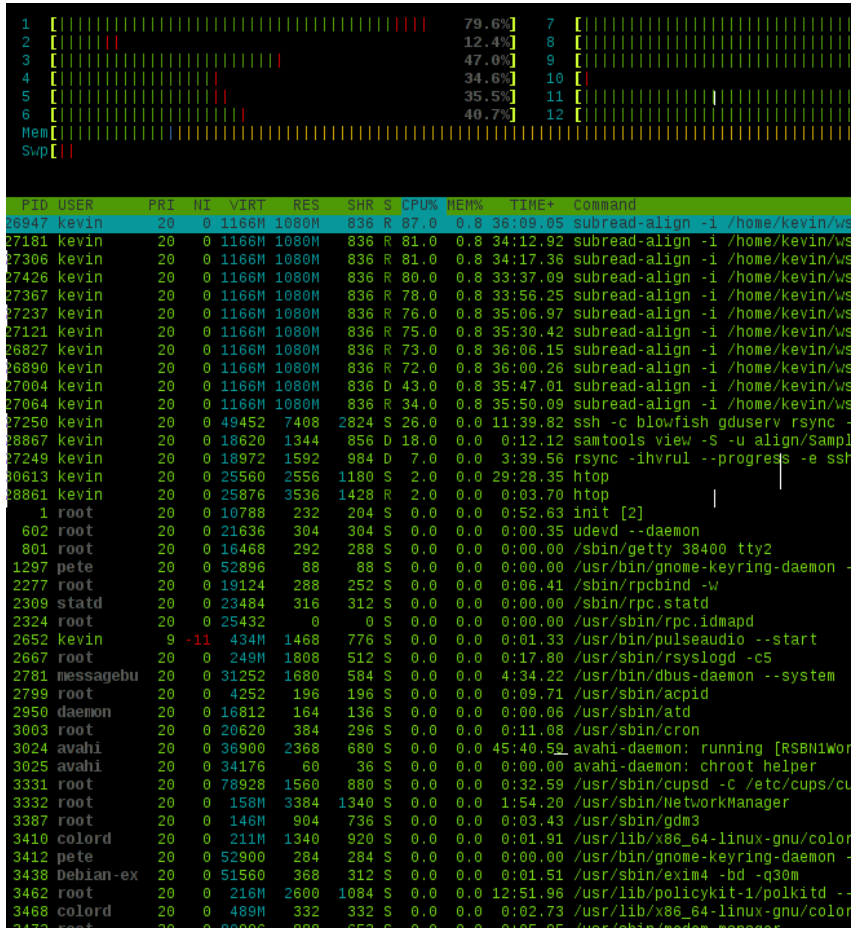



- Will focus on two areas of improvement
  - Making RNAseq library prep. cheaper & higher throughput
  - Making RNAseq data analysis easier & faster

# The Cazzonelli Button

*"Can't there just be a 'do my bioinformatics' button?"*

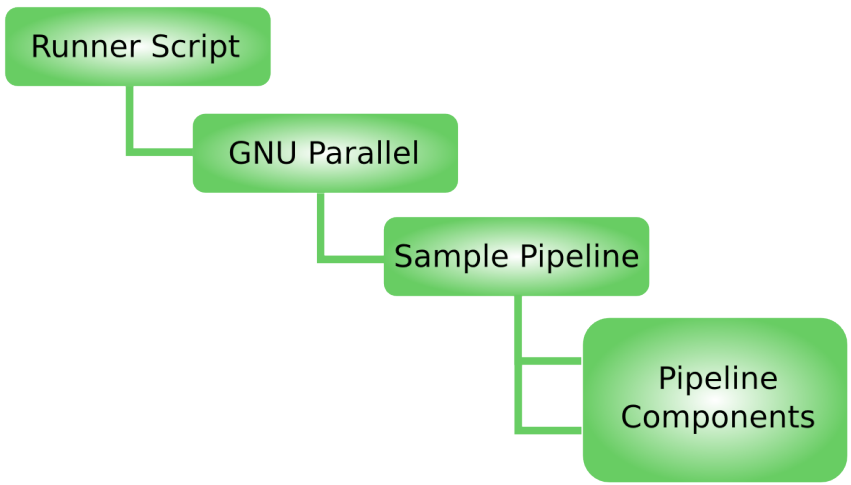
*Well, we laughed at you Chris, but its now* **almost** *that easy!*



# How To Run a Pipeline

## Command

```
bash runner.sh keyfile.key
```

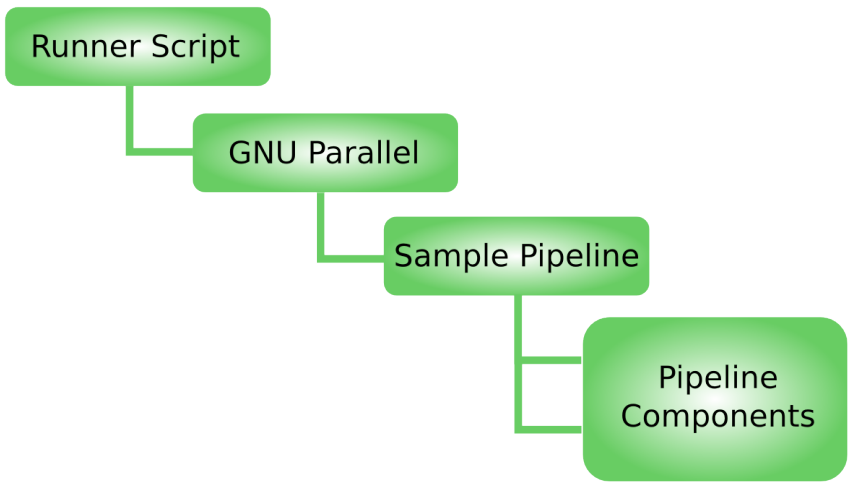



- Let's dissect that:
  - `bash runner.sh`

    Call the runner script
  - `keyfile.key`

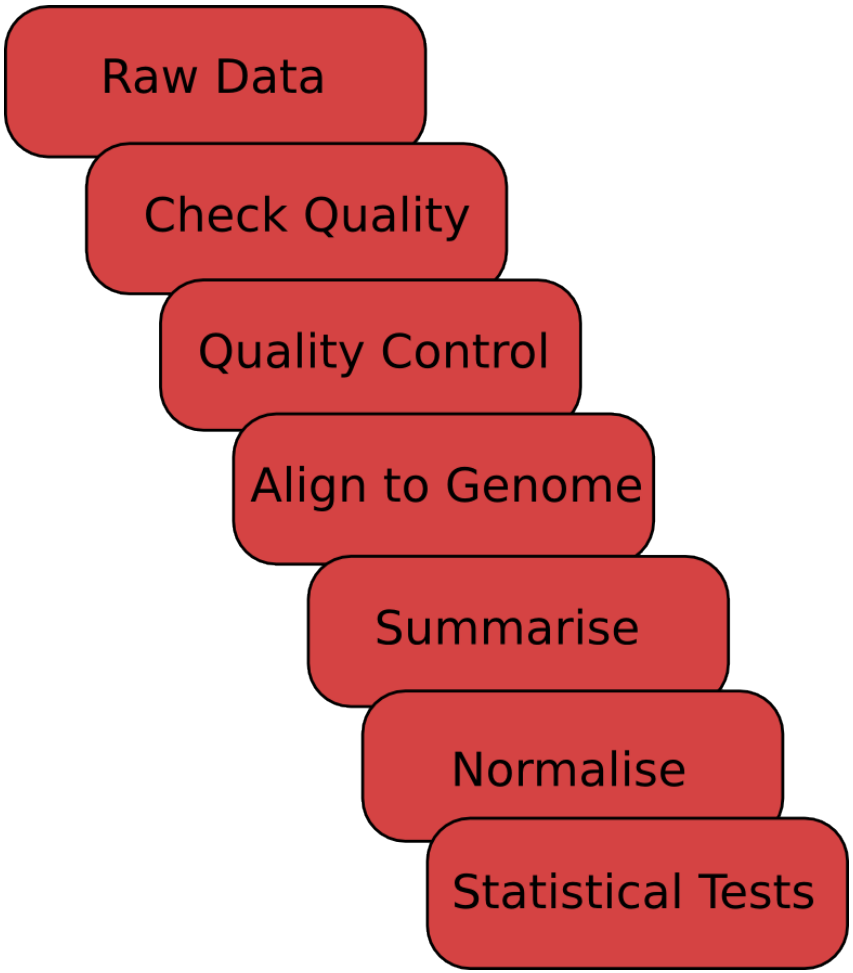
    Give it the "keyfile"
- No need to run each component separately

# How does that work?




- More than 1300 lines of code
- Written in bash, python and R
- 144 minor versions
- Code is on `github.com`
- Licensed under GNU GPL v3

# RNAseq Pipeline Components



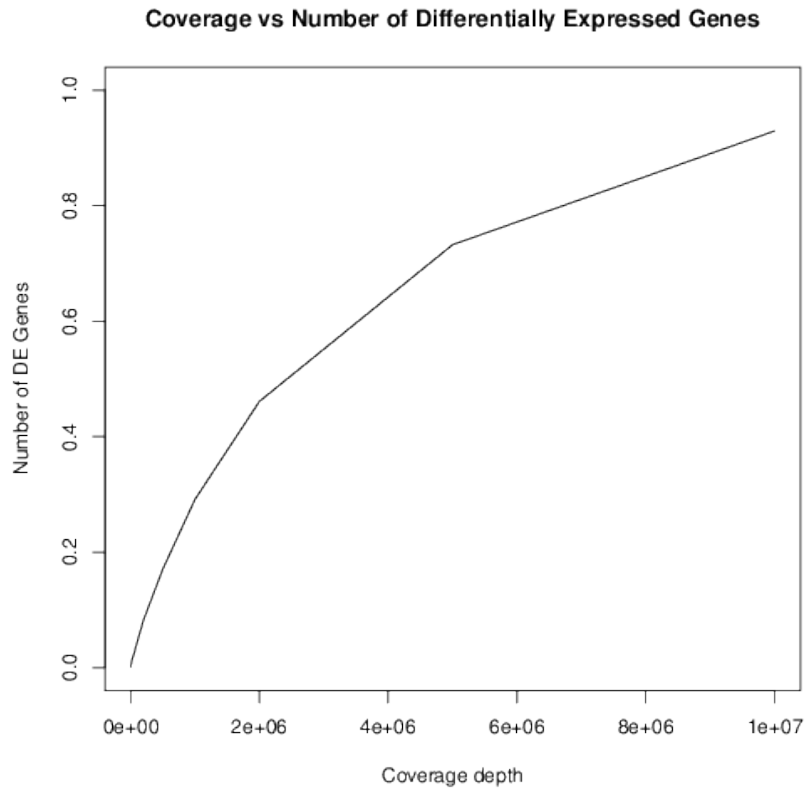

1. `fastqc`
2. `scythe`, `seqtk`
3. `subread` (`tophat`, `subjunc`)
4. `featureCounts`
5. `edgeR` (`DESeq`, `cuffdiff`)

# Effect of Coverage

- Coverage depth a limiting factor in RNAseq
- Aimed to empirically determine "cutoff" depth for our study system
- Hypothesised that 2 million reads gives 20% reduction in power
- 2 million reads gives $\approx$ 50% reduction
- Conclusion: use $>$ 2M reads per lane

# Effect of Coverage

- Coverage depth a limiting factor in RNAseq
- Aimed to empirically determine "cutoff" depth for our study system
- Hypothesised that 2 million reads gives 20% reduction in power
- 2 million reads gives $\approx$ 50% reduction
- Conclusion: use $>$ 2M reads per lane



# Aims

- Design & implement dynamic growth conditions
- Develop a pipeline of software to analyse RNAseq datasets
- 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

# QTL mapping aims

- Look for loci which control transcriptonal response to abiotic stress

# Growth of plants

- A QTL mapping set, parental lines and genetic controls
- 3 replicates
- Grown for 2 weeks, then dynamic growth conditions
- Assay expression before and after high light pulse treatment at 5 weeks

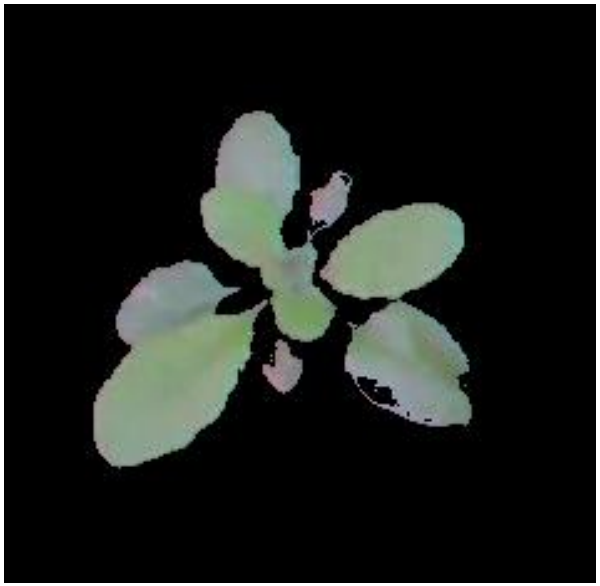
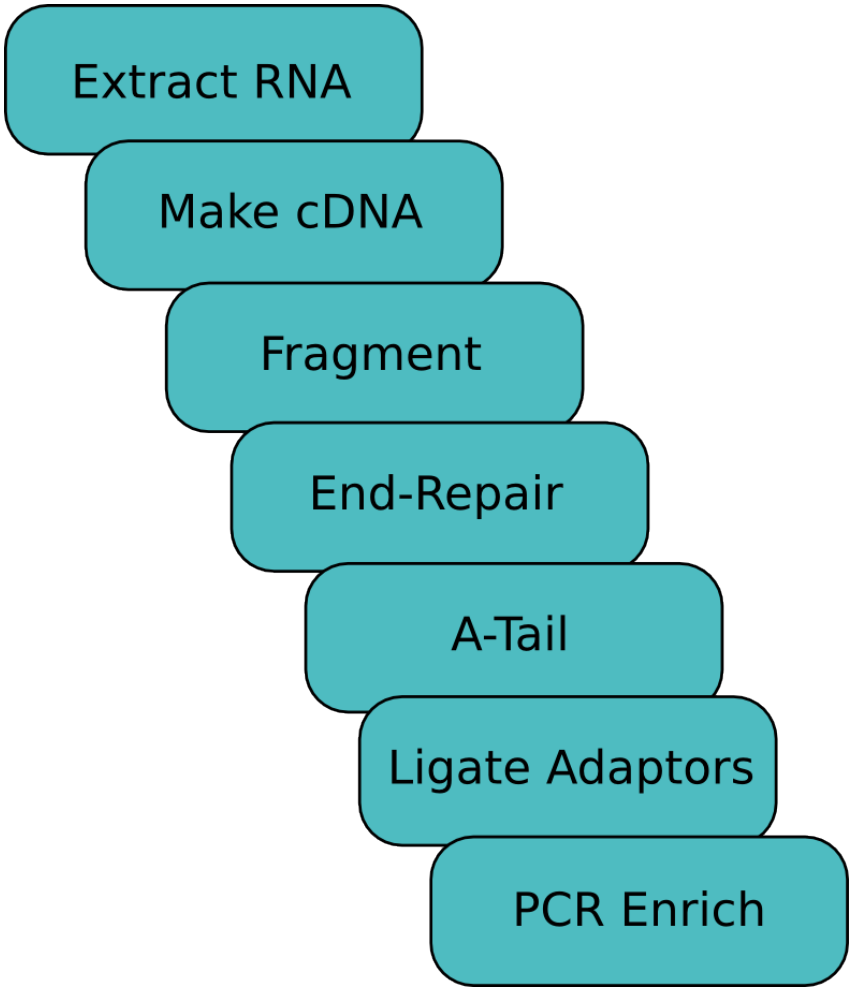


Figure: Sufficient Light

Figure: Excess Light

# Cheaper, Higher Throughput Method Needed

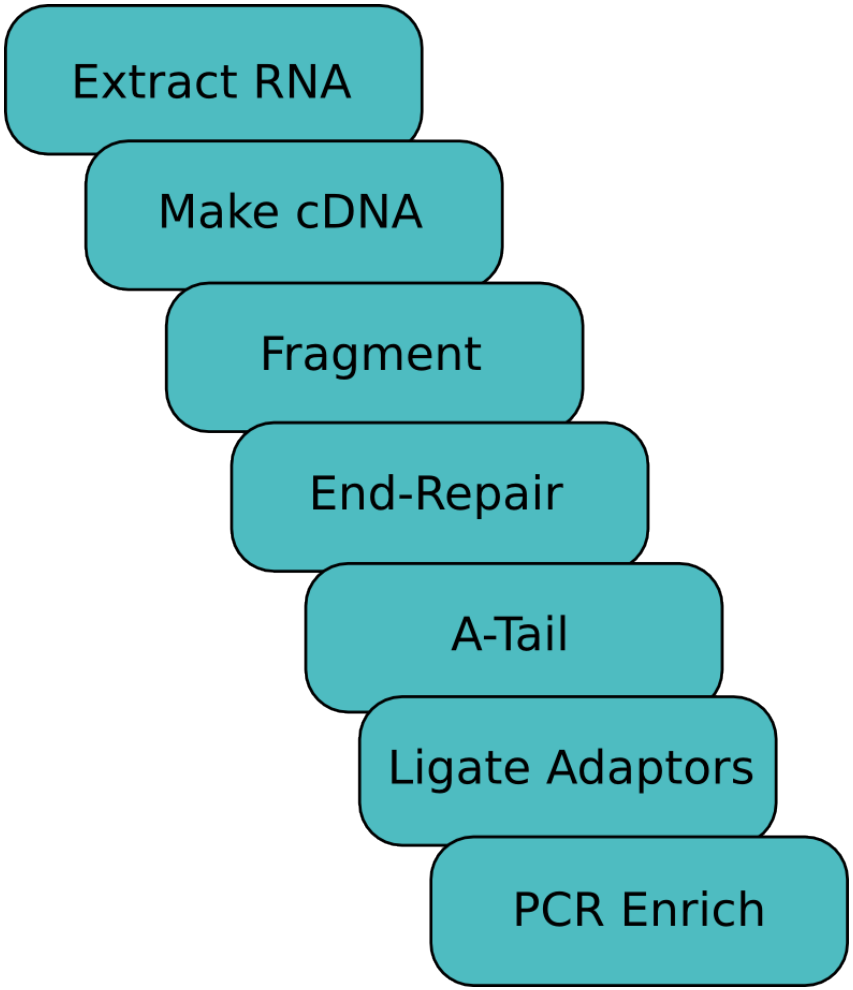



- Adapted from Kumar et al. (2012)
- On-bead SPRI protocol
- Performed in 96 well plate
- $\approx$ \$50 per sample, 96 samples per lane
- Successful until final step
- Sidelined due to lack of time

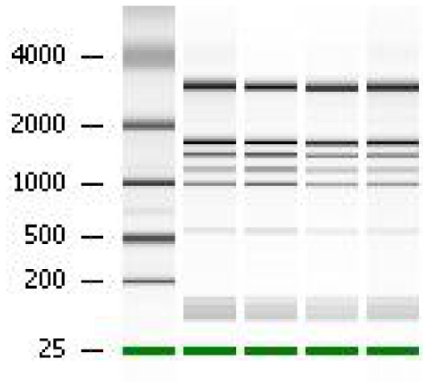
# Aims

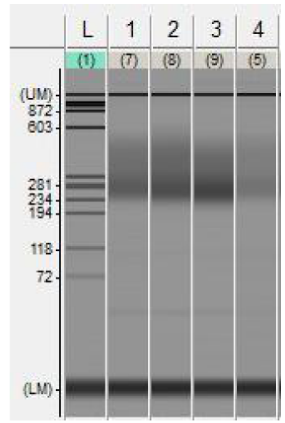
- Design & implement dynamic growth conditions
- Develop a pipeline of software to analyse RNAseq datasets
- Generate phenomic and transcriptomic QTL mapping datasets from plants grown under dynamic light conditions
- Determine effect of light intensity on transcriptome under dynamic light conditions

# Illumina RNAseq Library Prep.



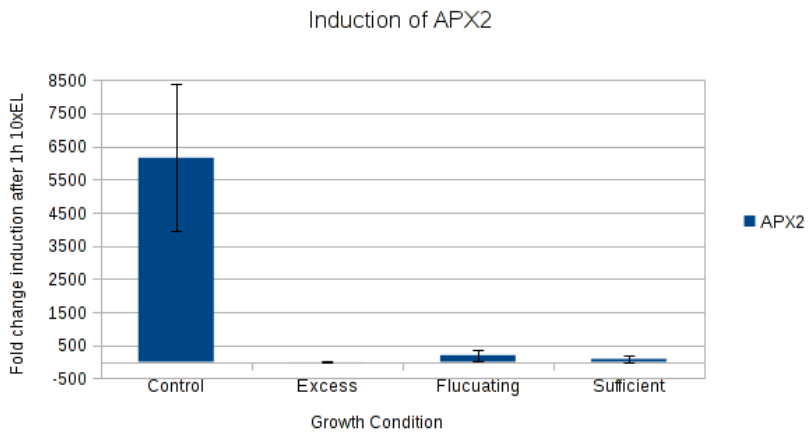

- 3-5 day protocol
- Up to 12 samples per lane
- $240-400 per sample



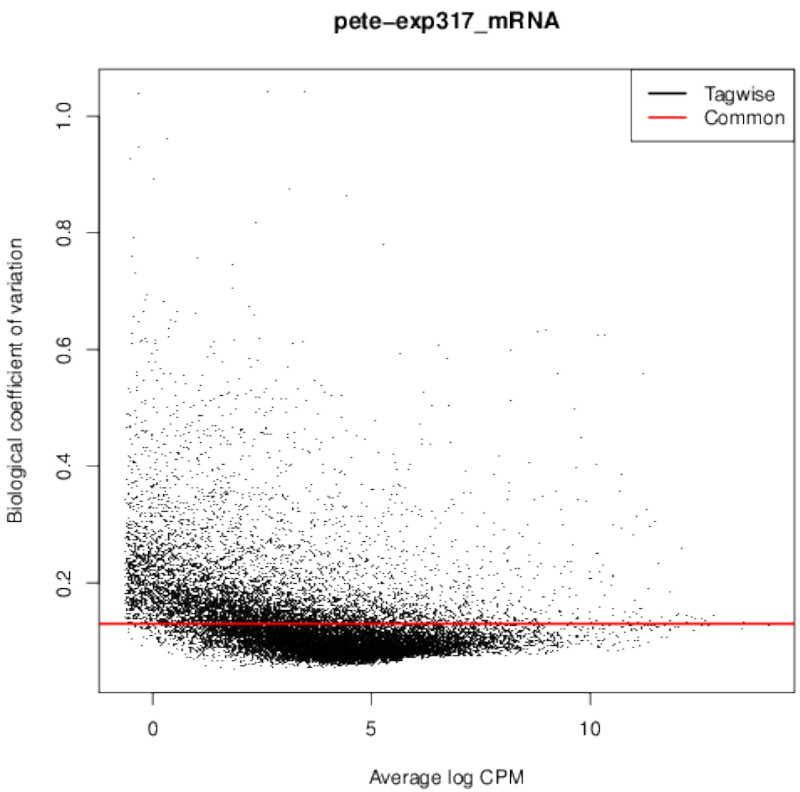
# qPCR analysis tells a small story

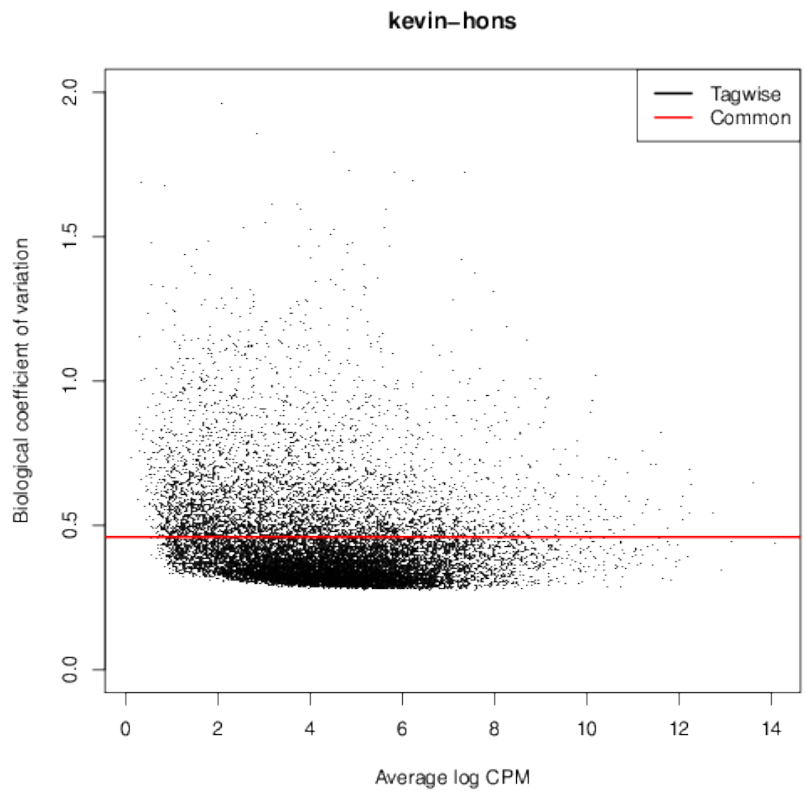
- Examine expression of known excess light responsive genes
- Show reduced induction and increased steady state expression
- Hypotheses appear mostly correct



# RNAseq shows the whole picture

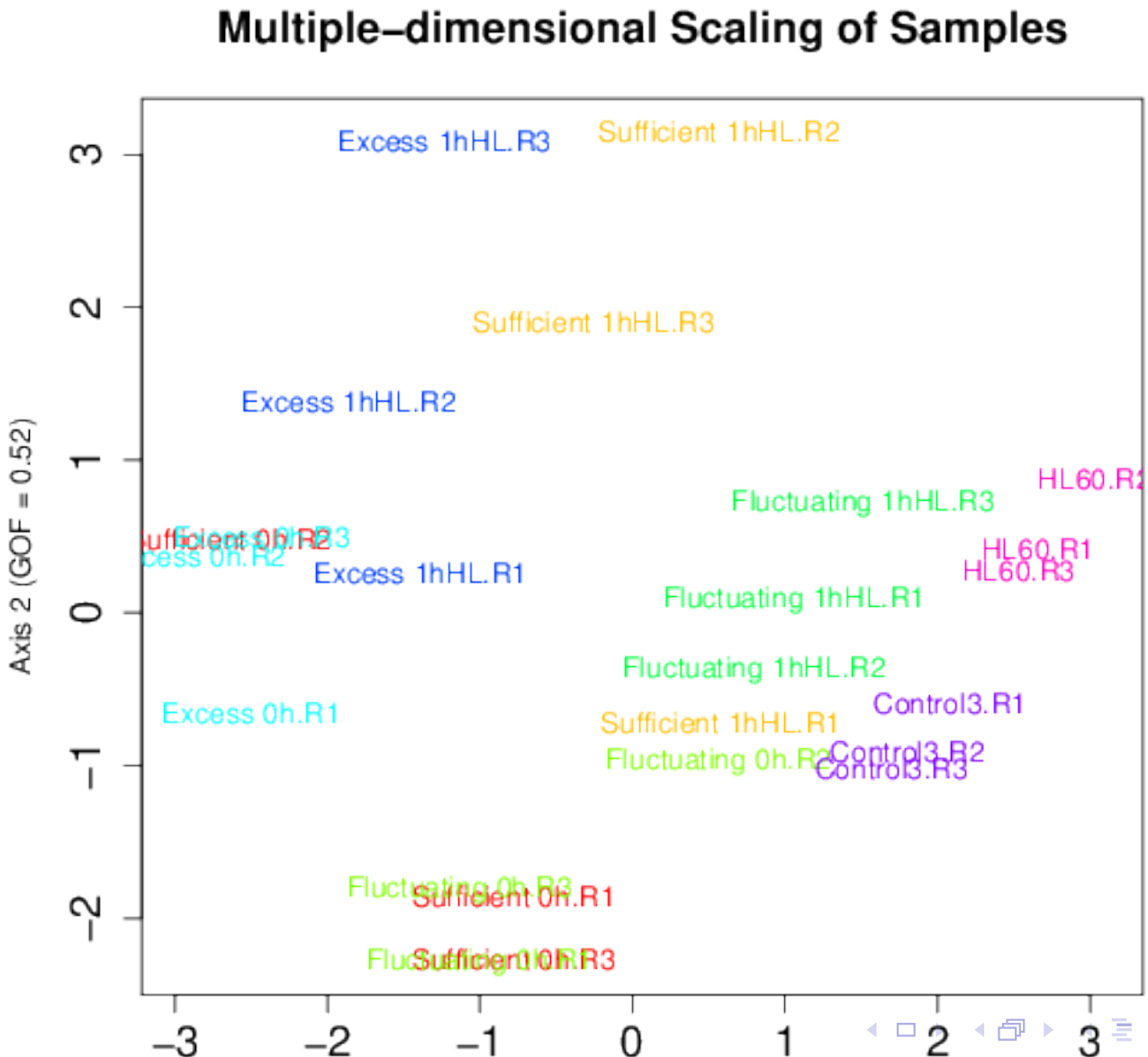
- Overall, high variation amongst samples



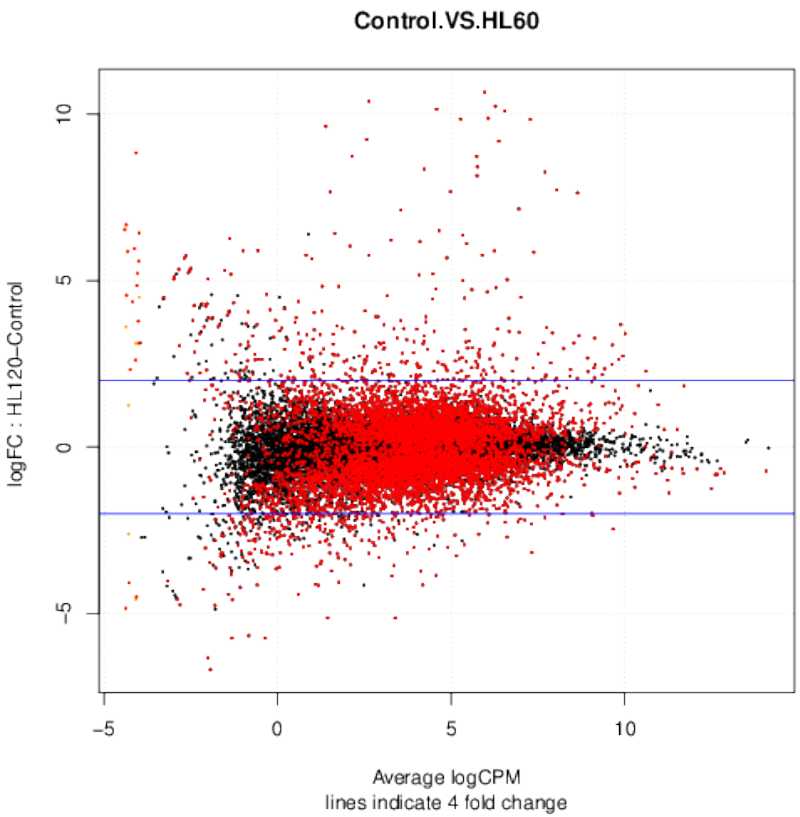
# RNAseq shows the whole picture

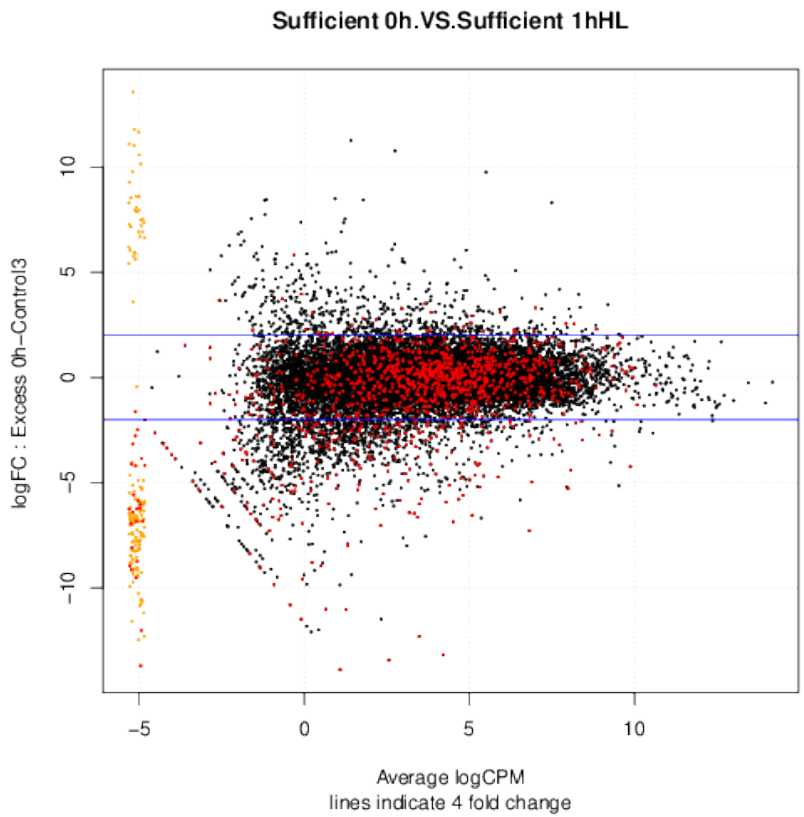
- Amongst a noisy response, a pattern emerges



# RNAseq shows the whole picture

- Less differential expression than expected



# DE table

- This is how many genes are DE
- This is what they have in common
- Analysis/interpretation is ongoing - (GSEA)

# Conclusions

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

# Future Work

- Optimise 96-well RNAseq protocol
- Analyse QTL mapping set:
  - RNAseq to map QTLs for expression of stress responsive genes
  - Find phenomic traits e.g. anthocyanin acclumulation
- Repeat entire experiment with improved sampling techniques - increase statsistical power

# Old/Dead slides

# 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

# Light Quantity

- Plants need a "happy medium" of light
- Too little = sub-optimal growth
- Too much = photooxidative damage
- Interesting model system for examining dynamic growth conditions

# Light Response & Transcriptomics

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

# History of Light Transcriptomics

- Pogson, Nagashi, Karpinski, Jannsen

# 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

# Making a pipeline?

+

# Making a pipeline?

+

# 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

# Steps in an RNAseq Analysis

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

# Steps in an RNAseq Analysis

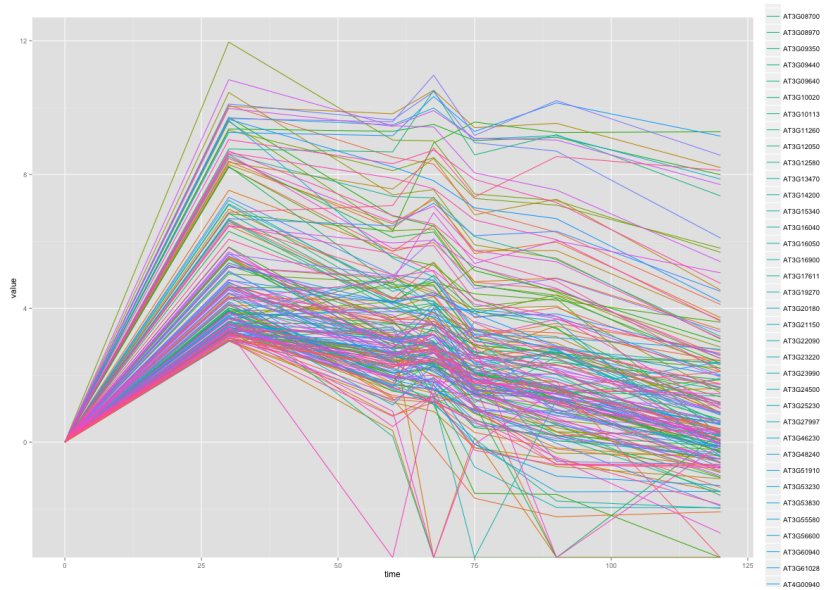
- Raw sequence data needs to be:
  - Filtered for quality
  - Aligned to the genome
- Once aligned, one needs to:
  - Quantify gene expression
  - Normalise counts
  - Perform statistical tests

# Existing "Best Practice" Pipeline

- Settings optimised for non plants
- Not optimised for large/dramatic changes
- Result in hypothesised weird artefacts
- Can result in false positives or false negatives

# Existing "Best Practice" Pipeline

- Settings optimised for non plants
- Not optimised for large/dramatic changes
- Result in hypothesised weird artefacts
- Can result in false positives or false negatives

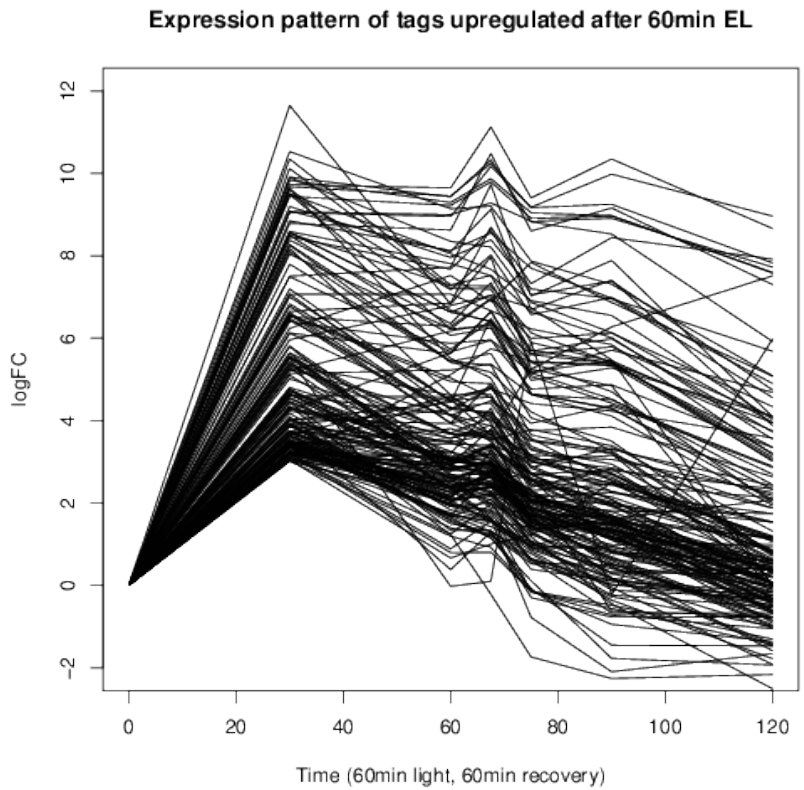


# "Improved" plot

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

# "Improved" plot

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



# Venn Diagrams

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