

Transcriptome Variation in *Arabidopsis* Under Dynamic Growth Conditions

or: How I learned to stop worrying and love RNAseq

Kevin Murray

Borevitz and Pogson Labs

September 25, 2013

Project Background

- ▶ Plants experience abiotic stress in nature
- ▶ Plants exhibit natural variation in stress tolerance
- ▶ Studying natural variation can give clues to mechanism
- ▶ Matching natural variation to dynamic conditions may uncover cryptic phenotypes



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

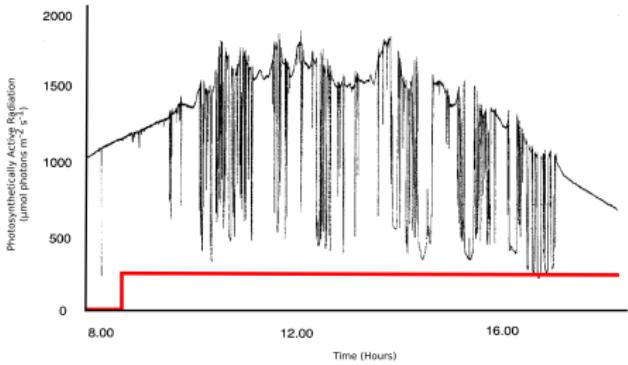
Aims

1. Design & implement dynamic growth conditions
2. Develop improved bioinformatic and molecular protocols for High-throughput RNAseq experiments
3. Determine effect of light intensity on transcriptome under dynamic light conditions

Aim 1: The growth condition dilemma



- ▶ Plants grow in nature
- ▶ A lot of science done in labs



(Külheim, Ågren, and Jansson 2002)

- ▶ Aim to merge elements of these two scenarios

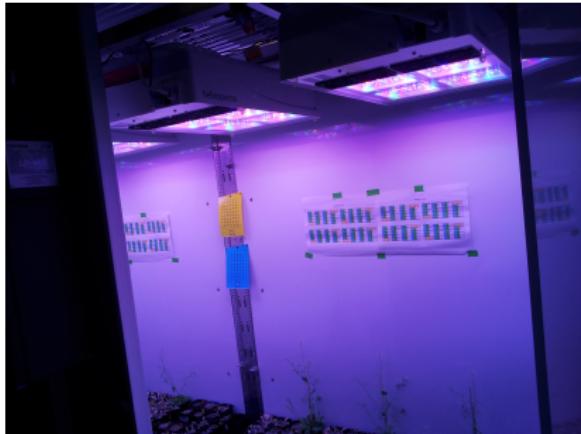
Aim 1: Introducing the SpectralPhenoClimatron

- ▶ Several new technologies
 - ▶ Growth Cabinets
 - ▶ LED Arrays
 - ▶ Imaging hardware
- ▶ Simulate regional climates
- ▶ Model diurnal and circannual trends of climate
- ▶ Use model simulation to drive actual growth cabinet conditions



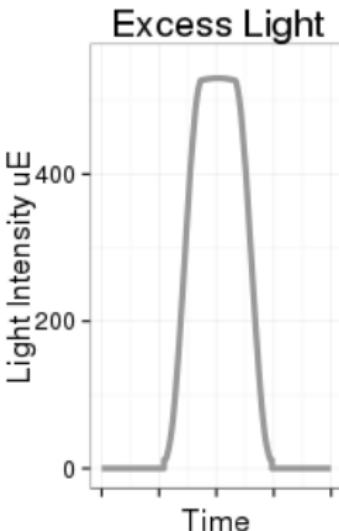
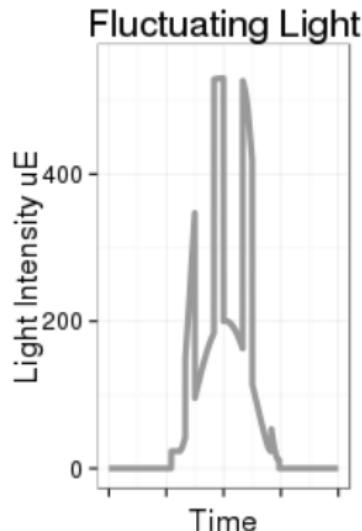
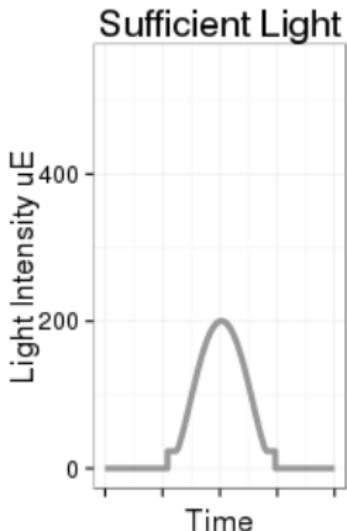
Aim 1: Result: Controlling the SpectralPhenoClimatron

- ▶ Disparate pieces of technology
- ▶ Need software “glue” to stick bits together
- ▶ Wrote `spcControl` Python module
 - ▶ 750 lines
 - ▶ 134 minor, 16 major versions
 - ▶ Open source, on github.com



Aim 1: Result: Novel Growth Conditions

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



Aim 1: Hypothesis: Plants Become Hardened

- ▶ 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

Aim 1: Plant Growth Under Dynamic Growth Conditions

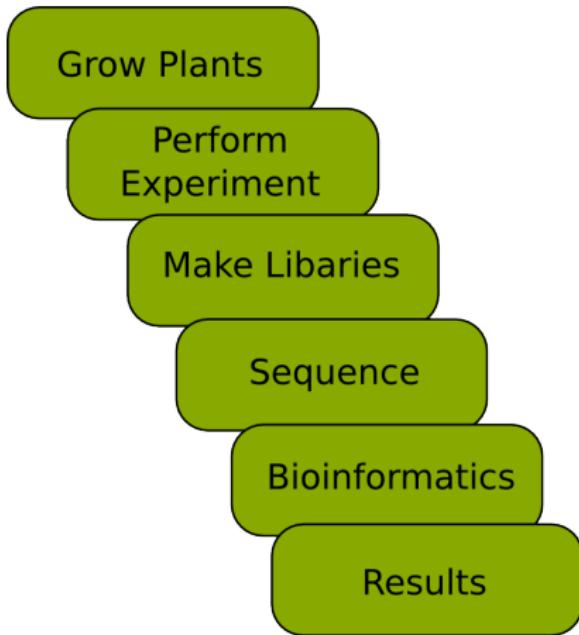
- ▶ A QTL mapping set; Col, Cvi, Ler Ecotypes
- ▶ Over 1200 plants planted
- ▶ Grown for 3 weeks dynamic growth conditions
- ▶ Assay expression before and after high light pulse treatment



Aims

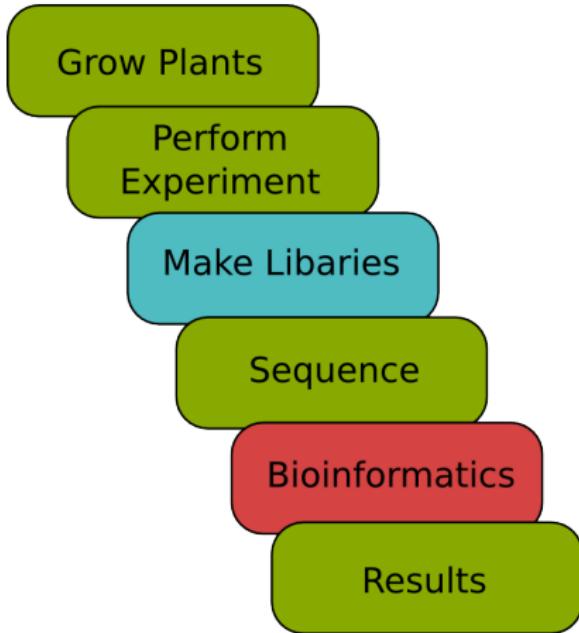
1. Design & implement dynamic growth conditions
2. Develop improved bioinformatic and molecular protocols for High-throughput RNAseq experiments
3. Determine effect of light intensity on transcriptome under dynamic light conditions

Aim 2: How does RNAseq work?



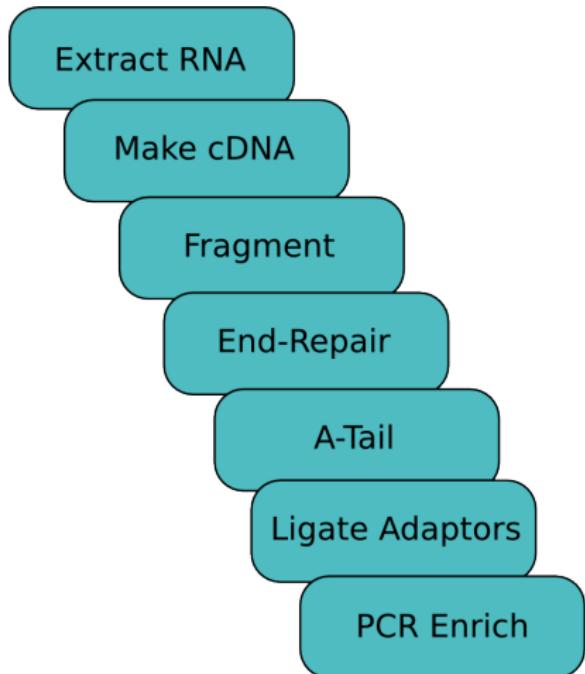
- ▶ Assay **ALL** expression in your tissue
- ▶ Unbiased, as quantitative as qPCR
- ▶ Becoming cheaper and easier

Aim 2: How does RNAseq work?



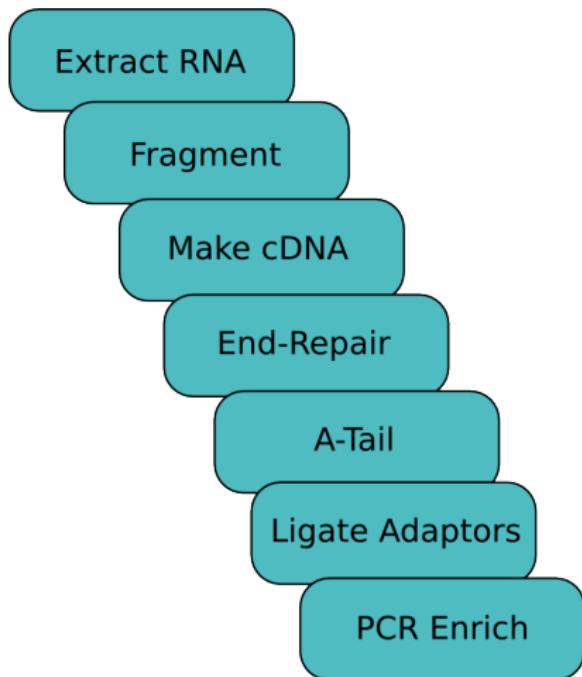
- ▶ Will focus on two areas of improvement
 - ▶ Making RNAseq library prep. cheaper & higher throughput
 - ▶ Making RNAseq data analysis easier & faster

Aim 2: Cheaper, Higher Throughput RNAseq

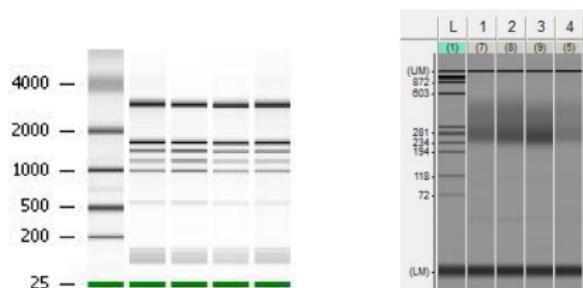


- ▶ Adapted from Kumar et al. (2012)
- ▶ On-bead SPRI protocol
- ▶ Performed in 96 well plate
- ▶ $\approx \$50$ per sample, 96 samples per lane
- ▶ Successful until final step
- ▶ Sidelined due to lack of time

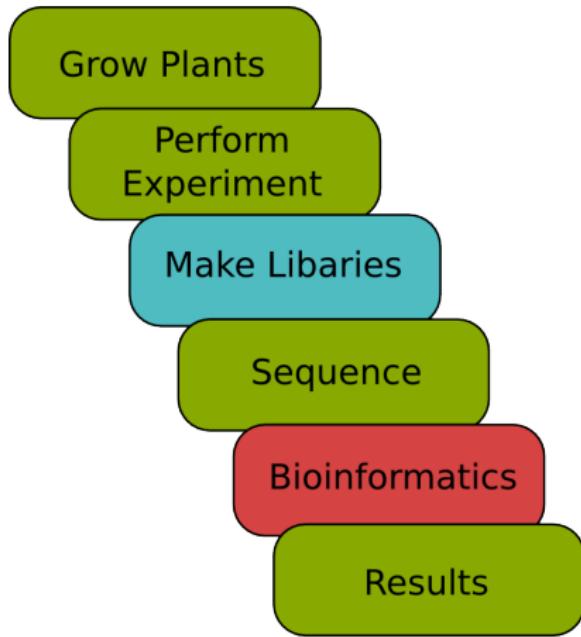
Aim 2: Illumina RNAseq Library Prep.



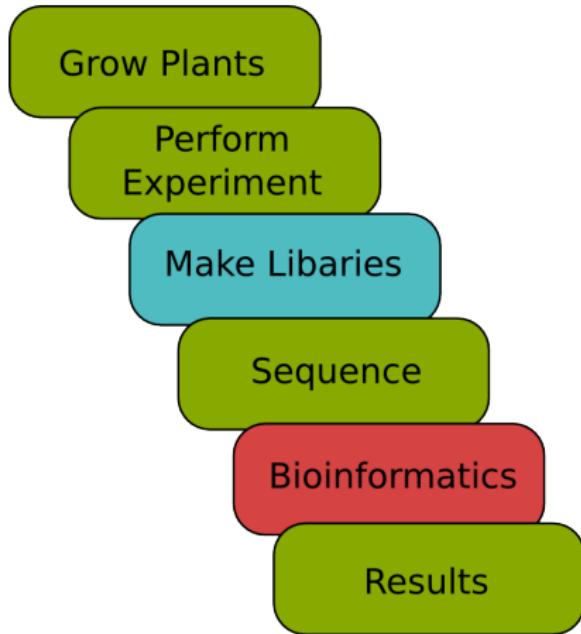
- ▶ 3-5 day protocol
- ▶ Up to 12 samples per lane
- ▶ \$240-400 per sample



Aim 2: RNAseq Analysis Made Easy!

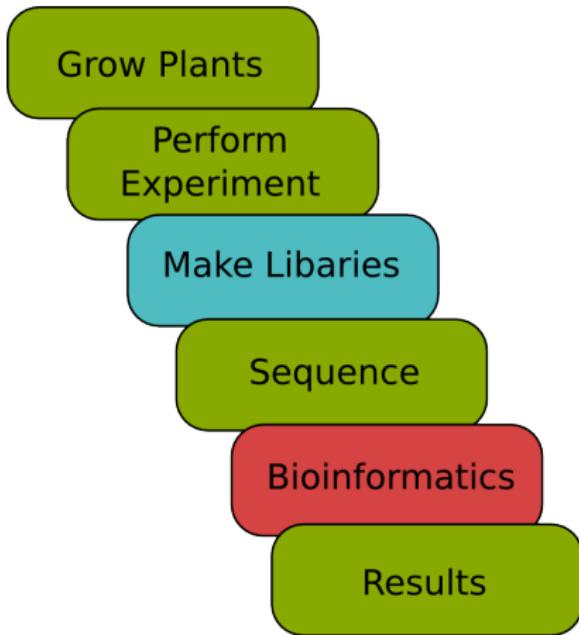


Aim 2: RNAseq Analysis Made Easy!



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

Aim 2: RNAseq Analysis Made Easy!



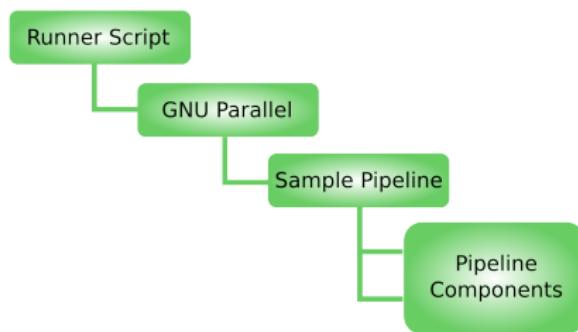
*"Can't there just be a
'do my bioinformatics'
button?"*



Aim 2: How To Run a Pipeline

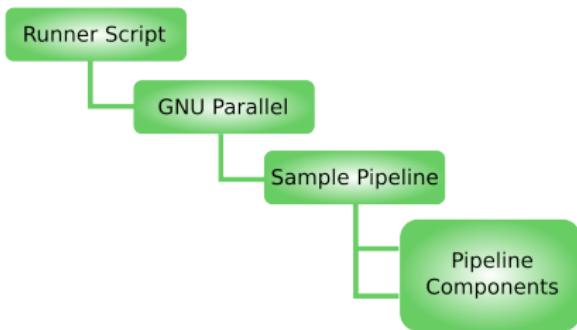
Command:

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



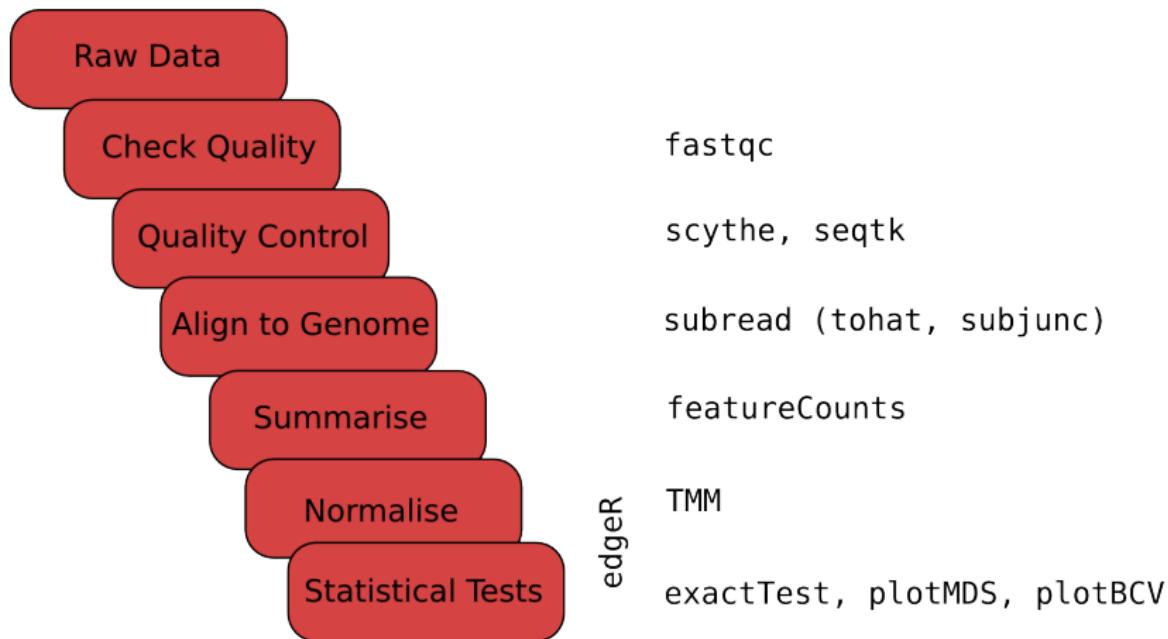
- ▶ Let's dissect that:
 - ▶ bash runner.sh
 - Call the runner script
 - ▶ keyfile.key
 - Give it the “keyfile”
- ▶ No need to run each component separately

Aim 2: How does that work?



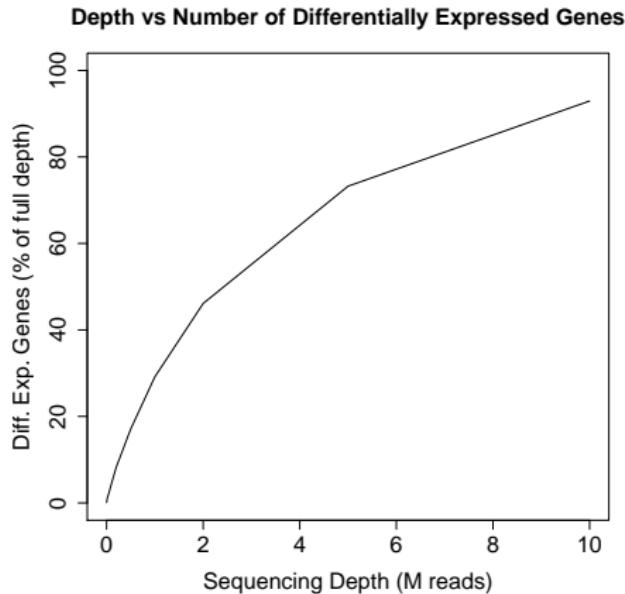
- ▶ More than 1300 lines of code
- ▶ Written in bash, python and R
- ▶ 144 minor versions
- ▶ Code is on github.com
- ▶ Open source (GPL v3)
- ▶ You should use it!

Aim 2: RNAseq Pipeline Components



Aim 2: Effect of Sequencing Depth

- ▶ Trade off between multiplexing and statistical power
- ▶ Conclusion: Recommend 48x multiplexing (5M reads)



Aims

1. Design & implement dynamic growth conditions
2. Develop improved bioinformatic and molecular protocols for High-throughput RNAseq experiments
3. Determine effect of light intensity on transcriptome under dynamic light conditions

Aim 3: Overview

Implement Novel
Growth Conditions

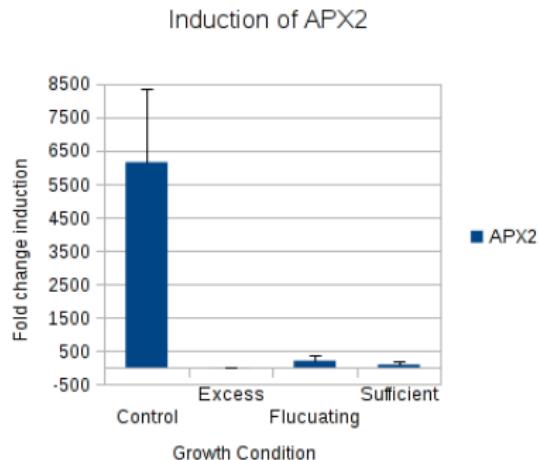
Grow 1200 Plants

Make Col-0 Libraries

Bioinformatics

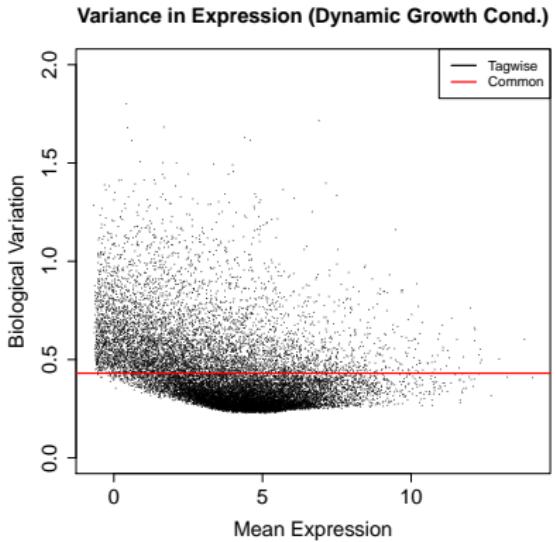
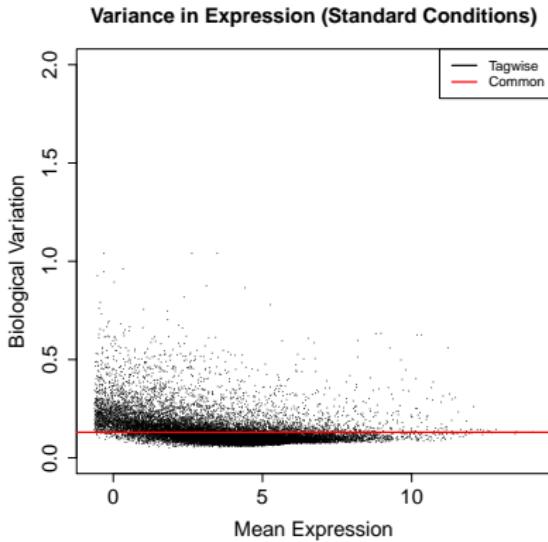
Aim 3: qPCR analysis tells a small story

- ▶ Examine expression of known excess light responsive genes (APX2, ELIP1, ELIP2, LHCBI.4)
- ▶ Dynamic growth conditions show reduced induction and increased steady state expression
- ▶ Hypotheses appear mostly correct



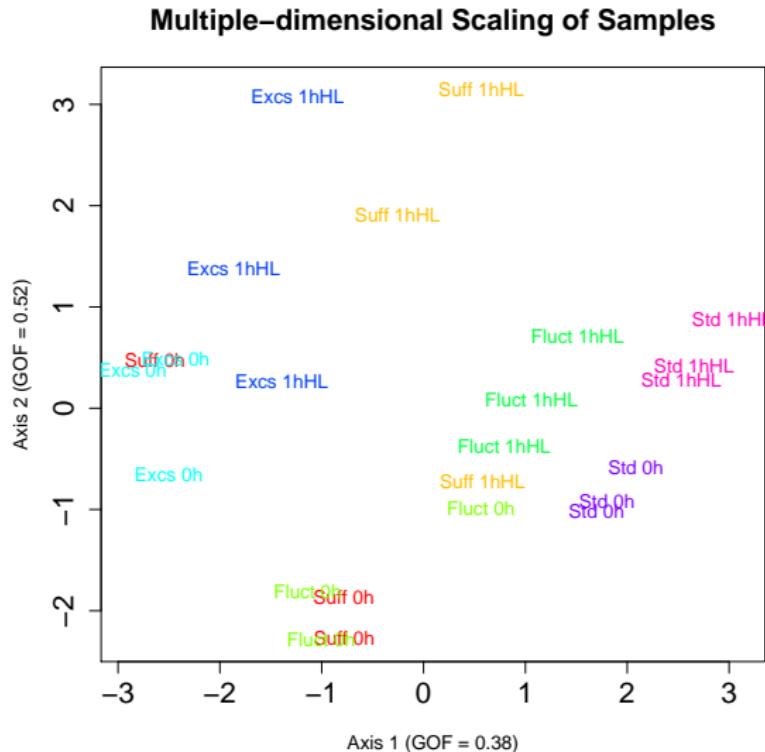
RNAseq shows the whole picture

- ▶ Overall, high variation amongst samples



RNAseq shows the whole picture

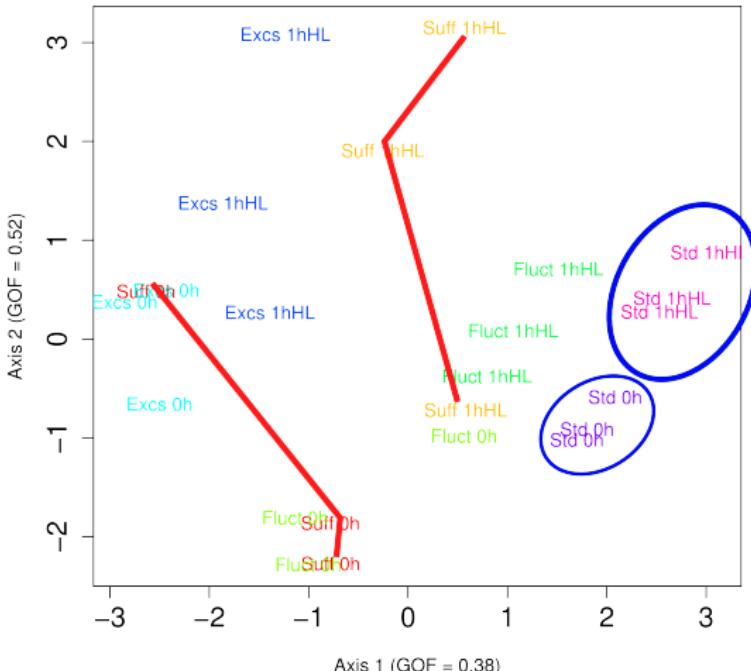
- Amongst a noisy response, a pattern emerges



RNAseq shows the whole picture

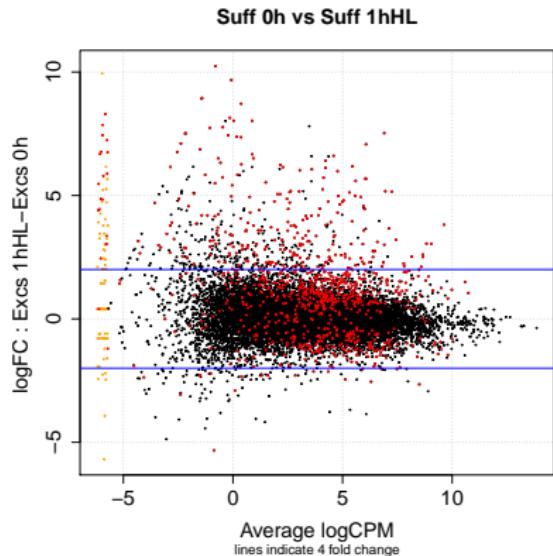
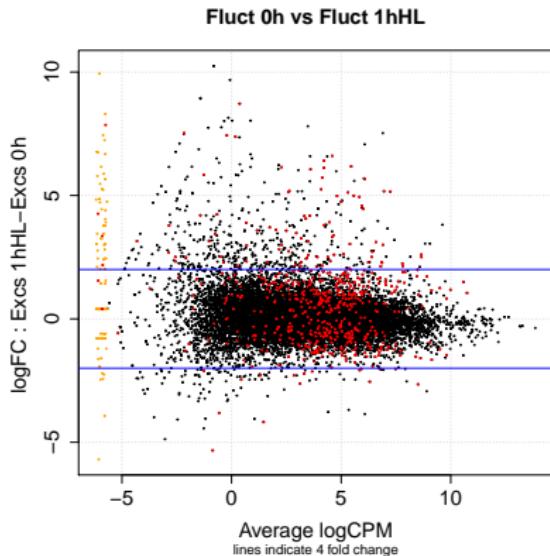
- ▶ Amongst a noisy response, a pattern emerges
- ▶ Plants grown in dynamic condition still exhibit stress response

Multiple-dimensional Scaling of Samples



RNAseq shows the whole picture

- Less differential expression in more “hardened” conditions



- 343 up, 229 down

- 666 up, 227 down

Conclusions

1. Design & implement dynamic growth conditions
 - ▶ *Dynamic growth conditions may allow stress “hardening”*
2. Develop improved bioinformatic and molecular protocols for High-throughput RNAseq experiments
 - ▶ *RNAseq is here, and easier than you might think*
3. Determine effect of light intensity on transcriptome under dynamic light conditions
 - ▶ *Patterns of differential expression seen, more replicates needed, analysis ongoing*

Future Work

- ▶ Optimise 96-well RNAseq protocol - required for expression QTL mapping
- ▶ Analyse QTL mapping set:
 - ▶ RNAseq to map expression QTLs of stress responsive genes
 - ▶ Map QTLs for phenomic traits e.g. anthocyanin accumulation
 - preliminary data looks interesting
- ▶ Repeat entire experiment with improved sampling techniques - increase statistical power

Acknowledgements

- ▶ Pogson Lab
- ▶ Borevitz lab
- ▶ Special thanks to Pete Crisp and Norman Warthmann

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

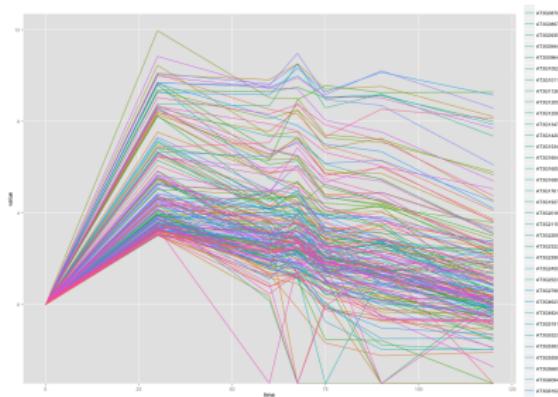
- ▶ Raw sequence data needs to be:
 - ▶ Filtered for quality
 - ▶ Aligned to the genome
- ▶ Once aligned, one needs to:
 - ▶ Quantify gene expression
 - ▶ Normalise counts
 - ▶ Perform statistical tests

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

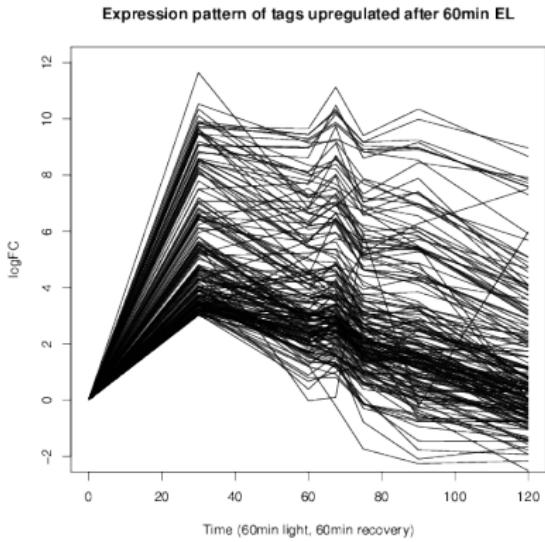


“Improved” plot

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

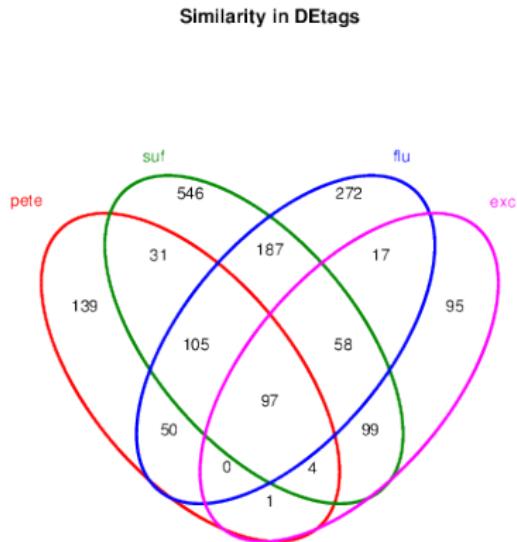
“Improved” plot

- ▶ 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



Unique objects: All = 1701; S1 = 427; S2 = 1127; S3 = 786; S4 = 371

QTL mapping aims

- ▶ Look for loci which control transcriptional response to abiotic stress