

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?

1	[79.6%	
2]										7	
3											12.4%	
4											47.4%	
5											34.6%	
6											35.5%	
7											11	
8											12	
9											40.7%	
Me												
Sus												
TF2D	NAME	PRIO	NI	VIRT	RES	SHR	UPC/N	HEAVY	TIMER	COMMAND		
1721	kevin	20	0	1168M	1080M	835 R	81.0	0.8	34.12	0.2	subread-align -i /home/kevin/v3	
7308	kevin	20	0	1168M	1080M	835 R	81.0	0.8	34.17	3.6	subread-align -i /home/kevin/v3	
17426	kevin	20	0	1168M	1080M	835 R	80.0	0.8	33.37	0.9	subread-align -i /home/kevin/v3	
17367	kevin	20	0	1168M	1080M	835 R	78.0	0.8	33.56	2.5	subread-align -i /home/kevin/v3	
17237	kevin	20	0	1168M	1080M	835 R	76.0	0.8	35.86	0.7	subread-align -i /home/kevin/v3	
7121	kevin	20	0	1168M	1080M	835 R	75.0	0.8	35.30	4.2	subread-align -i /home/kevin/v3	
5827	kevin	20	0	1168M	1080M	835 R	73.0	0.8	36.06	0.15	subread-align -i /home/kevin/v3	
6890	kevin	20	0	1168M	1080M	835 R	72.0	0.8	36.00	0.25	subread-align -i /home/kevin/v3	
7204	kevin	20	0	1168M	1080M	835 R	43.0	0.8	35.47	0.81	subread-align -i /home/kevin/v3	
17238	kevin	20	0	1168M	1080M	835 R	42.0	0.8	35.47	0.81	subread-align -i /home/kevin/v3	
17258	kevin	20	0	49452	7408	2324 S	36.0	0.8	11.95	0.2	ssn -c blowfish primary rsyslog	
8867	kevin	20	0	10620	1344	856 D	18.0	0.8	0	12.12	sartools view -S u sign/samples	
7249	kevin	20	0	18972	592	884 D	7.0	0.8	3.39	56	rsync -i hrvul --progress -e ssh	
76613	kevin	20	0	25568	2556	1180 S	2.0	0.8	29.28	35.70	rsync -i hrvul --progress -e ssh	
8861	kevin	20	0	25876	3536	1428 R	2.0	0.8	0	83.70	httpd	
1 root		20	0	10788	232	204 S	0.0	0.8	0	52.63	init [2]	
602 root		20	0	21636	304	304 S	0.0	0.8	0	0.00	0.05	udevd -n daemon
801 root		20	0	16468	292	288 S	0.0	0.8	0	0.00	0.00	ssn/getty 38486 tty2
1297 pete		20	0	52809	88	88 S	0.0	0.8	0	0.00	0.00	usr/bin/gnome-keyring-daemon
2950 daniel		20	0	16000	292	295 S	0.0	0.8	0	0.00	0.00	ssn/getty 38487 tty3
3200 stidol		20	0	33484	314	312 S	0.0	0.8	0	0.00	0.00	ssn/getty 38488 tty4
3234 root		20	0	52430	42	0 S	0.0	0.8	0	0.00	0.00	usr/sbin/rpc_idmapd
2652 kevin		20	9	11	4248	1468	775 S	0.0	0.8	0	81.33	usr/bin/pulseaudio --start
2667 root		20	0	2489	1808	512 S	0.0	0.8	0	17.80	usr/sbin/syslogd -c5	
2781 mesagabu		20	0	31252	1680	584 S	0.0	0.8	0	43.22	usr/bin/dbus-daemon --system	
2799 root		20	0	425	196	196 S	0.0	0.8	0	0.09	71	usr/sbin/acpid
2950 daniel		20	0	16812	164	136 S	0.0	0.8	0	0.00	0.00	ssn/getty 38489 tty5
3003 root		20	0	26026	384	296 S	0.0	0.8	0	0.11	0.00	ssn/getty 38490 tty6
3024 avahid		20	0	38900	2368	680 S	0.0	0.8	0	45.40	53.95	avahi-daemon: running [helper]
3131 root		20	0	78928	1580	880 S	0.0	0.8	0	0.00	0.00	ssn/getty 38491 tty7
3332 root		20	0	1588	3384	1340 S	0.0	0.8	0	15.20	0.40	usr/sbin/wheelmanager
3387 root		20	0	1486	904	736 S	0.0	0.8	0	0.03	0.43	usr/sbin/qdm3
3410 colorad		20	0	2138	1340	920 S	0.0	0.8	0	0.01	0.91	usr/lib/x86_64-linux-gnu/colorad
3412 pete		20	0	52988	284	284 S	0.0	0.8	0	0.00	0.00	usr/bin/gnome-keyring-daemon
3438 deban-ed		20	0	51569	368	312 S	0.0	0.8	0	0.01	0.51	usr/bin/eb -d -q39m
3462 root		20	0	2168	2600	1084 S	0.0	0.8	0	12.51	96	usr/lib/policykit-1/polkit
3468 colorad		20	0	4898	332	332 S	0.0	0.8	0	0.02	0.73	usr/lib/x86_64-linux-gnu/colorad

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



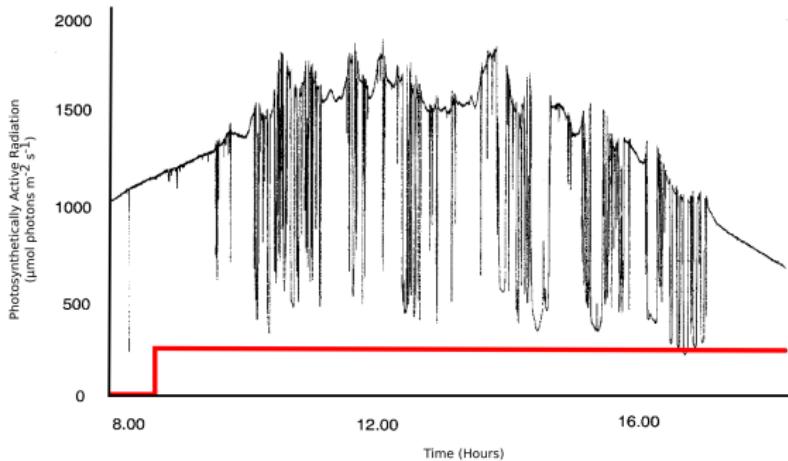
- ▶ Plants grow in nature (mostly)



- ▶ 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
- ▶ 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

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



Making a pipeline?



+



Making 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

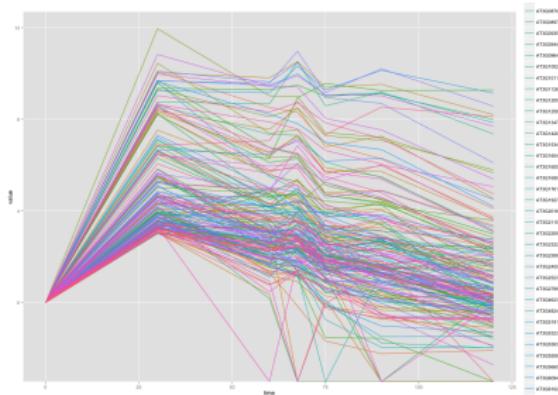
- ▶ Raw sequence data needs to be:
 - ▶ Filtered for quality
 - ▶ Aligned to the genome
- ▶ Once aligned, one needs to:
 - ▶ Quantify gene expression
 - ▶ Normalise counts
 - ▶ Perform statistical tests

Existing “Best Practice” Pipeline

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

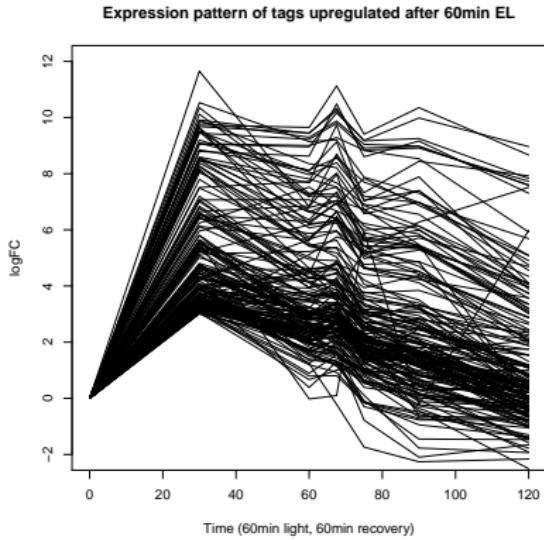


“Improved” plot

- ▶ 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

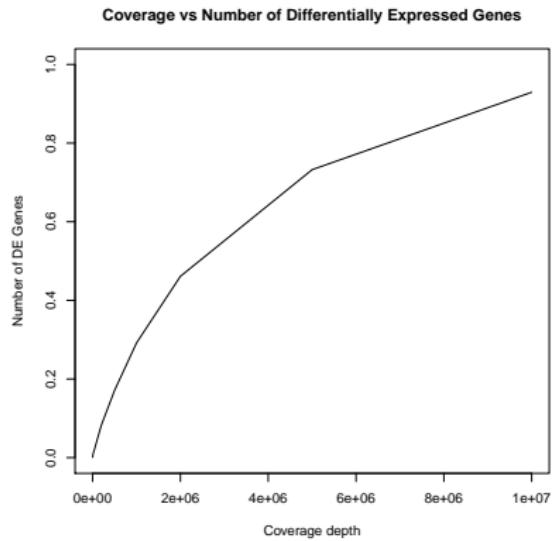


Effect of Coverage

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



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

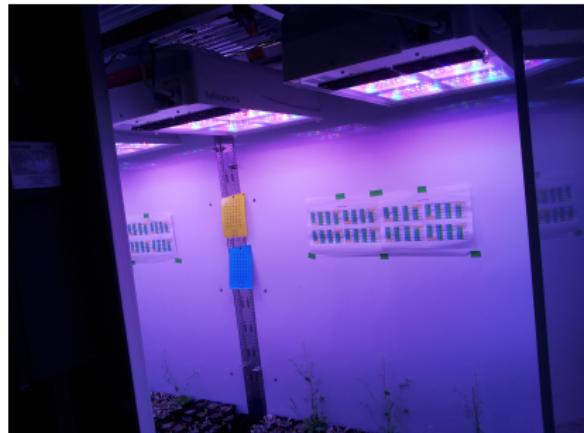
Aim 2

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

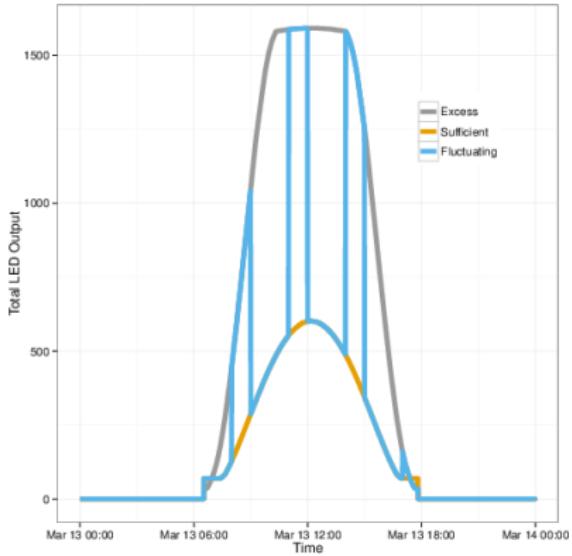
Introducing the SpectralPhenoClimatron

- ▶ 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

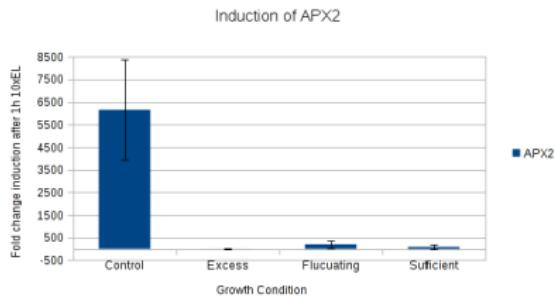


Investigating altered light intensity

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

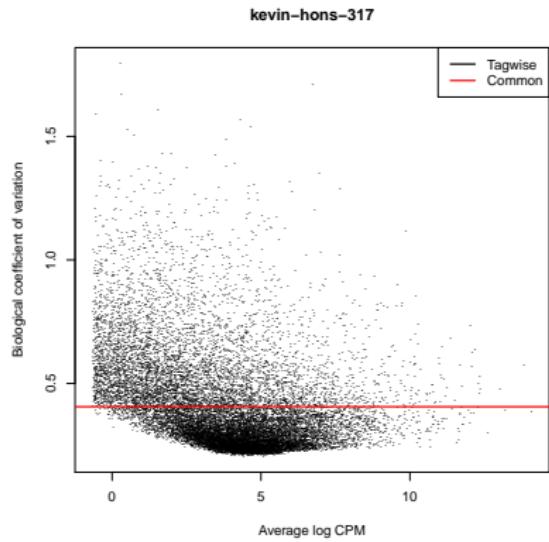
qPCR analysis tells a small story

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



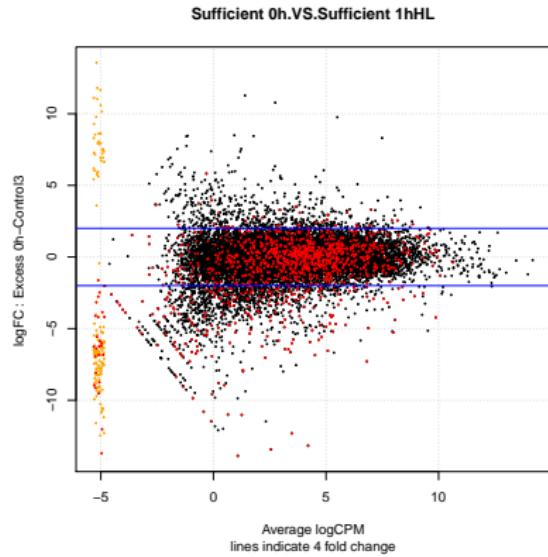
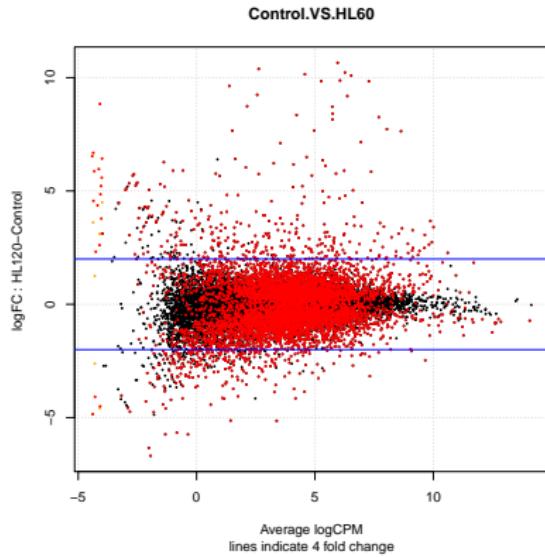
RNAseq shows the whole picture

- ▶ Overall, high variation amongst samples



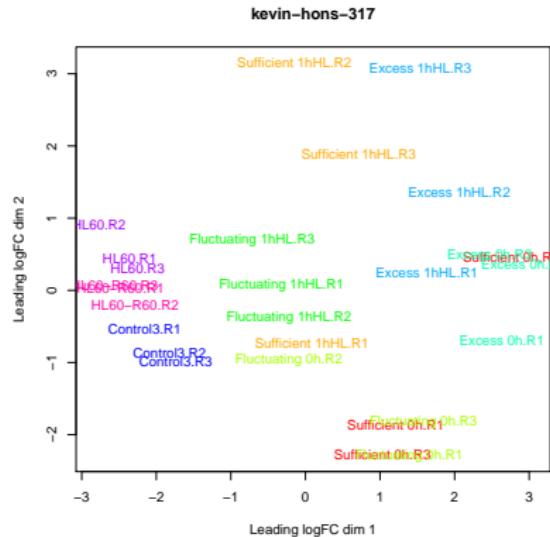
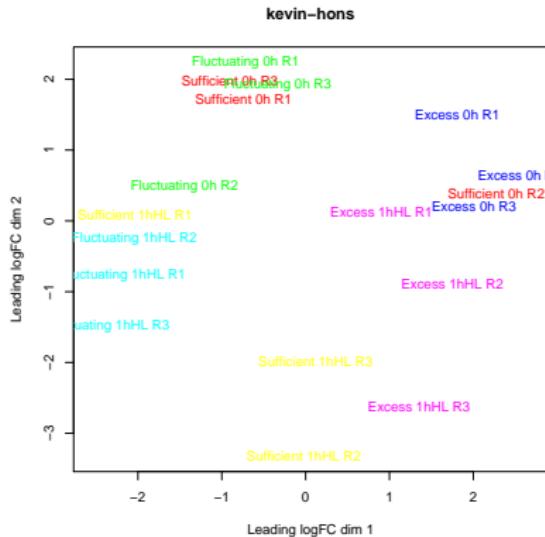
RNAseq shows the whole picture

- Less differential expression than expected



RNAseq shows the whole picture

- ▶ Amongst a noisy response, a pattern emerges



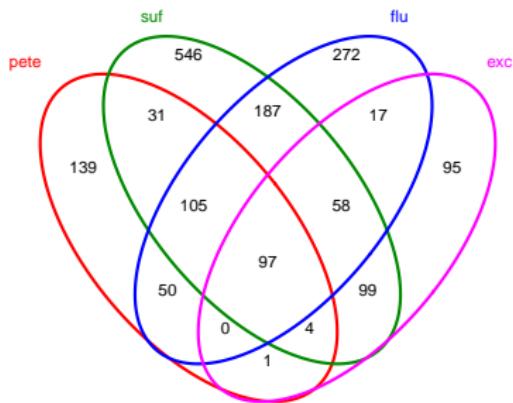
Conclusions

- ▶ 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



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

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

