Improved Statistical Methods for Analyzing Circadian Rhythms in High-Throughput Data

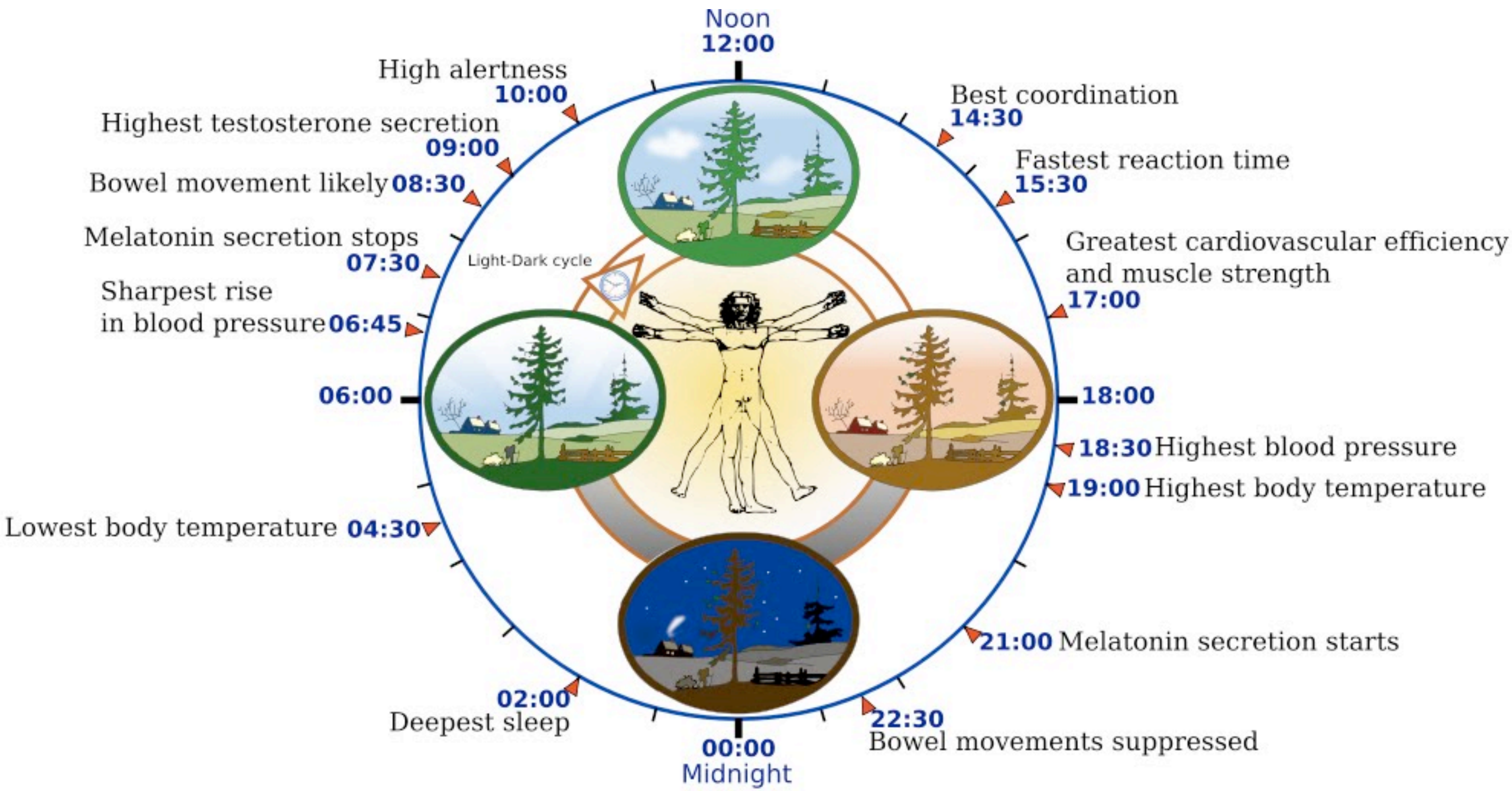
Alan L. Hutchison
Dinner Group

2016-08-26

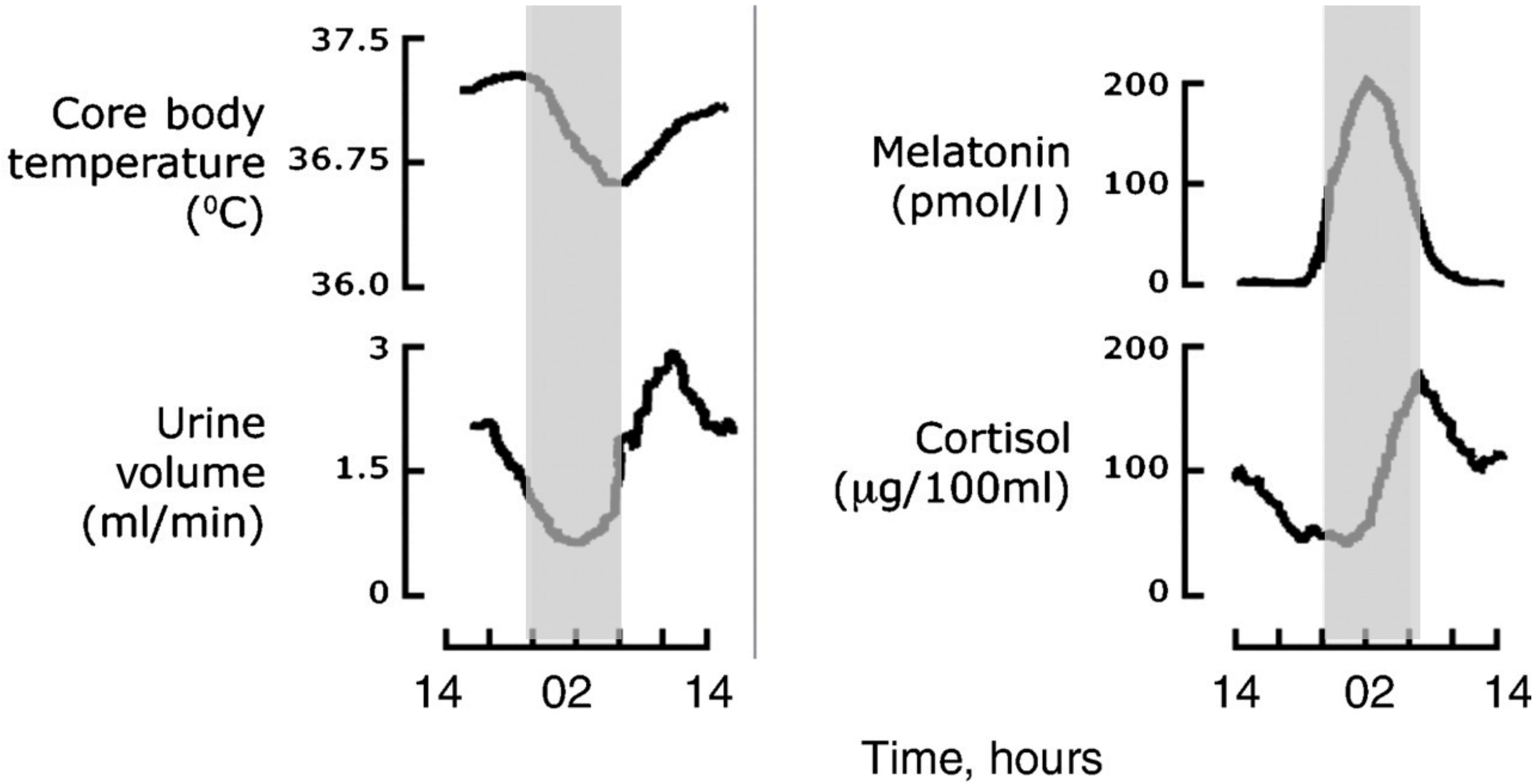
Photo courtesy of City of Chicago


LIFE

Circadian Rhythms are physiological rhythms regulated by an internal clock

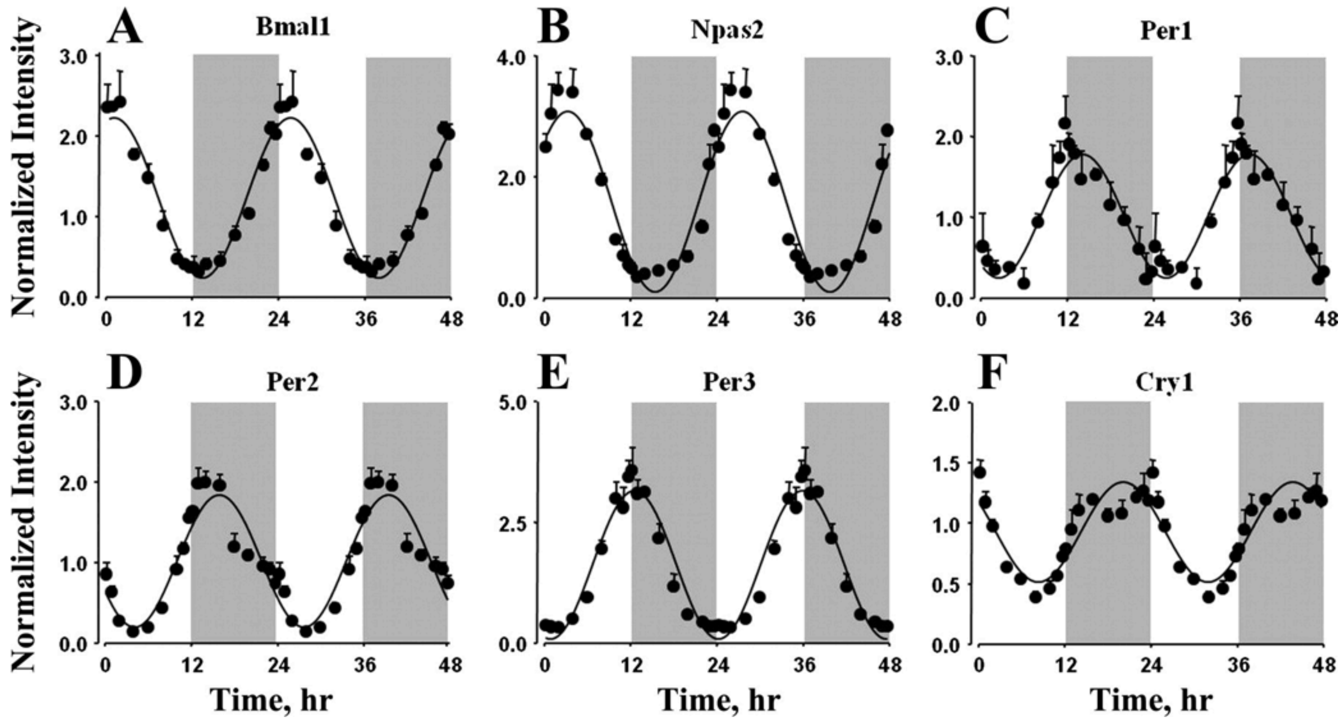


Many hormonal and physiological processes display circadian rhythms



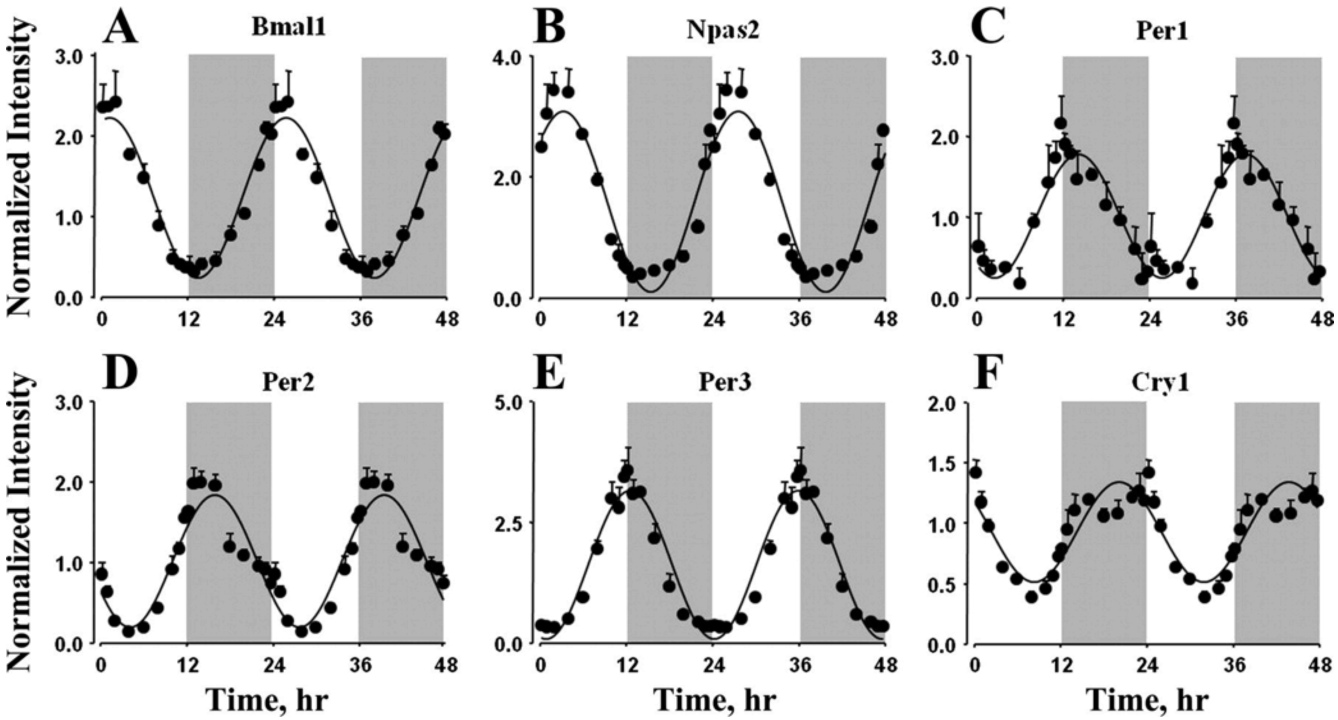
Hastings M, et al. J. Endocrinol. Nov 1, 2007. 195 187-198

A molecular transcriptional set of feedback loops controls circadian rhythms in eukaryotes



Sukumaran S. et al. Journal of Applied Physiology. 2011 Vol. 110 no. 6, 1732-1747

A molecular transcriptional set of feedback loops controls circadian rhythms in eukaryotes



Sukumaran S. et al. Journal of Applied Physiology. 2011 Vol. 110 no. 6, 1732-1747

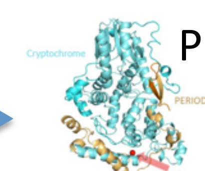
DNA

http://albany.bivfsites.com/wp-content/uploads/sites/25/2014/01/dna.png

RNA

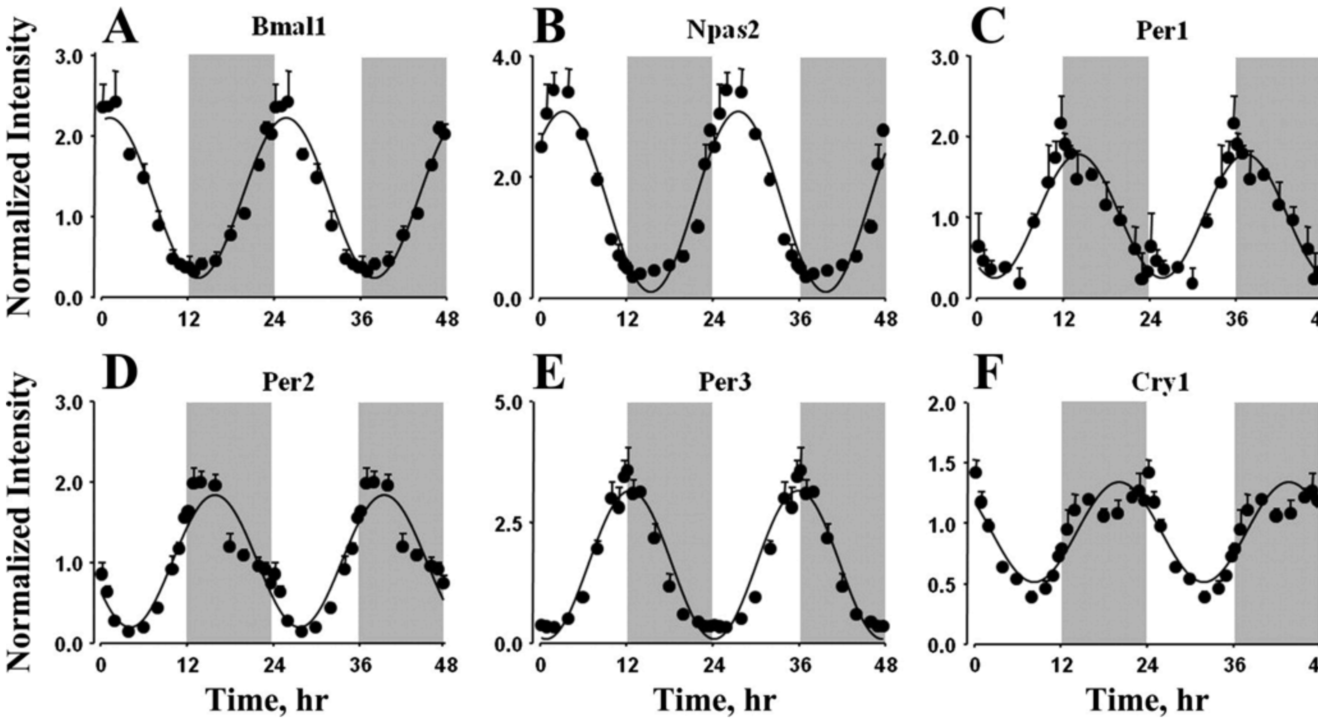
http://cdn2.hubspot.net/hub/127518/file-30478736-jpg/images/dna-rna-purification.jpg

Protein

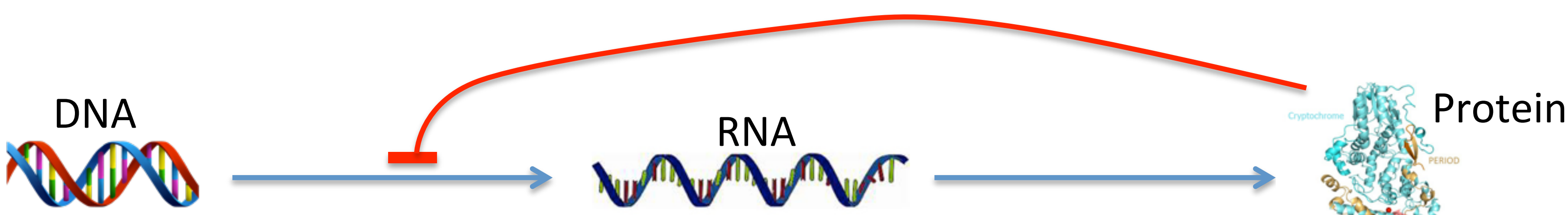


http://phys.org/news/2014-05-atomic-essential-circadian-clock-protein.html

A molecular transcriptional set of feedback loops controls circadian rhythms in eukaryotes



Sukumaran S. et al. Journal of Applied Physiology. 2011 Vol. 110 no. 6, 1732-1747

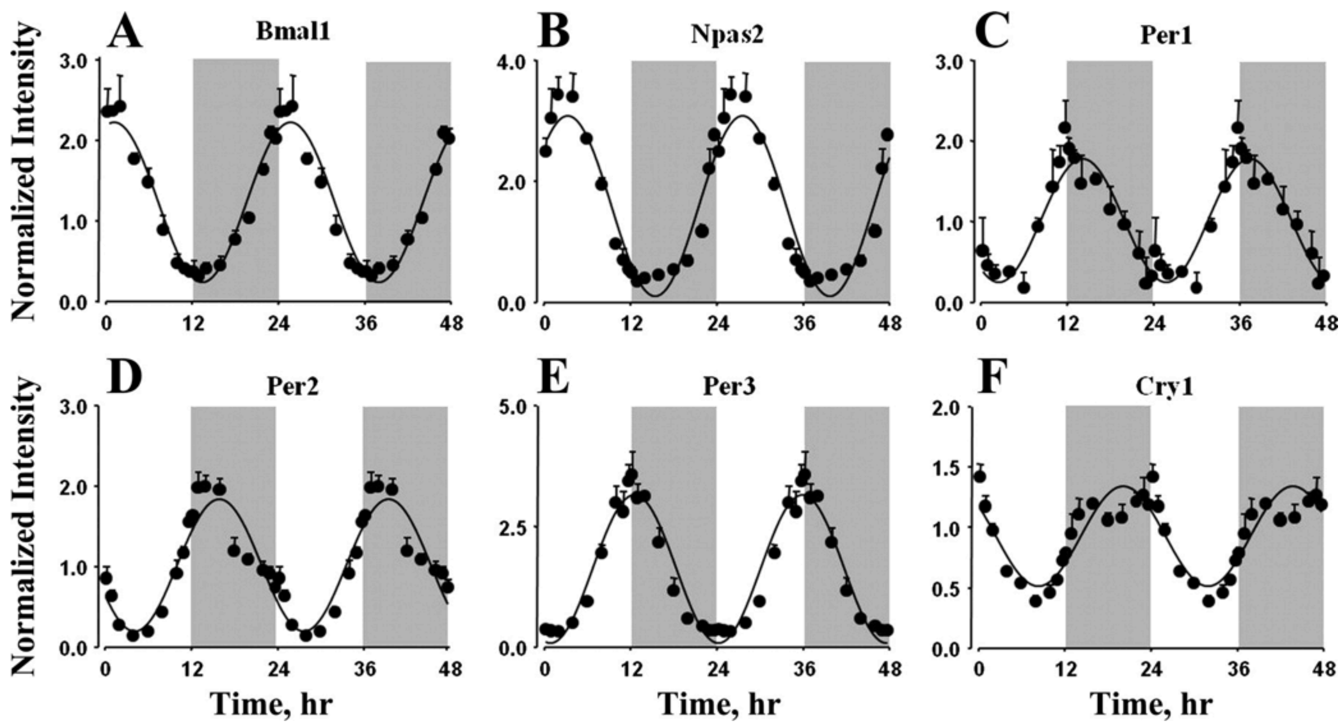


http://albany.bivfsites.com/wp-content/uploads/sites/25/2014/01/dna.png

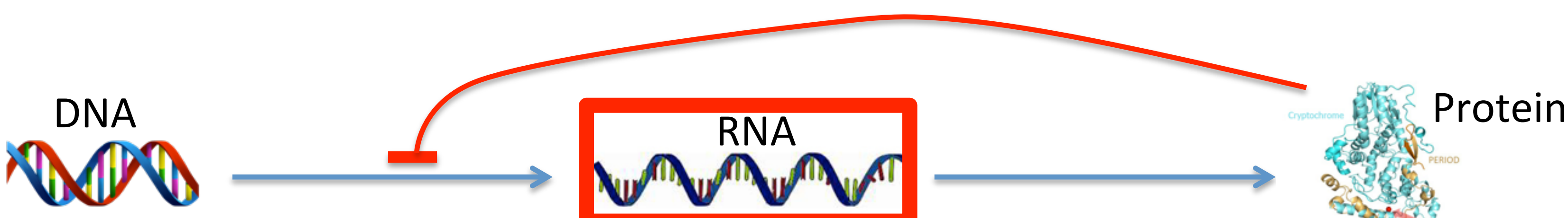
http://cdn2.hubspot.net/hub/127518/file-30478736-jpg/images/dna-rna-purification.jpg

http://phys.org/news/2014-05-atomic-essential-circadian-clock-protein.html

A molecular transcriptional set of feedback loops controls circadian rhythms in eukaryotes



Sukumaran S. et al. Journal of Applied Physiology. 2011 Vol. 110 no. 6, 1732-1747



http://albany.bivfsites.com/wp-content/uploads/sites/25/2014/01/dna.png

http://cdn2.hubspot.net/hub/127518/file-30478736-jpg/images/dna-rna-purification.jpg

http://phys.org/news/2014-05-atomic-essential-circadian-clock-protein.html

Medical relevance of Circadian Rhythms

Medical relevance of Circadian Rhythms

Adverse metabolic and cardiovascular consequences of circadian misalignment

Frank A. J. L. Scheer[a,b,1], Michael F. Hilton[a,2], Christos S. Mantzoros[b,c] and Steven A. Shea[a,b]

Medical relevance of Circadian Rhythms

Adverse metabolic and cardiovascular consequences of circadian misalignment

Frank A. J. L. Scheer[a,b,1], Michael F. Hilton[a,2], Christos S. Mantzoros[b,c] and Steven A. Shea[a,b]

Morning vaccination enhances antibody response over afternoon vaccination: A cluster-randomised trial

Joanna E. Long[a], Mark T. Drayson[b], Angela E. Taylor[d], Kai M. Toellner[b], Janet M. Lord[c, 1], Anna C. Phillips[a, 1,],

Medical relevance of Circadian Rhythms

Adverse metabolic and cardiovascular consequences of circadian misalignment

Frank A. J. L. Scheer[a,b,1], Michael F. Hilton[a,2], Christos S. Mantzoros[b,c] and Steven A. Shea[a,b]

Morning vaccination enhances antibody response over afternoon vaccination: A cluster-randomised trial

Joanna E. Long[a], Mark T. Drayson[b], Angela E. Taylor[d], Kai M. Toellner[b], Janet M. Lord[c, 1], Anna C. Phillips[a, 1]

Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes

Biliana Marcheva Kathryn Moynihan Ramsey Ethan D. Buhr Yumiko Kobayashi Hong Su Caroline H. Ko Ganka Ivanova Chiaki Omura Shelley Mo Martha H. Vitaterna James P. Lopez Louis H. Philipson Christopher A. Bradfield Seth D. Crosby Lellean JeBailey Xiaozhong Wang Joseph S. Takahashi Joseph Bass

Medical relevance of Circadian Rhythms

Adverse metabolic and cardiovascular consequences of circadian misalignment

Frank A. J. L. Scheer[a,b,1], Michael F. Hilton[a,2], Christos S. Mantzoros[b,c] and Steven A. Shea[a,b]

Morning vaccination enhances antibody response over afternoon vaccination: A cluster-randomised trial

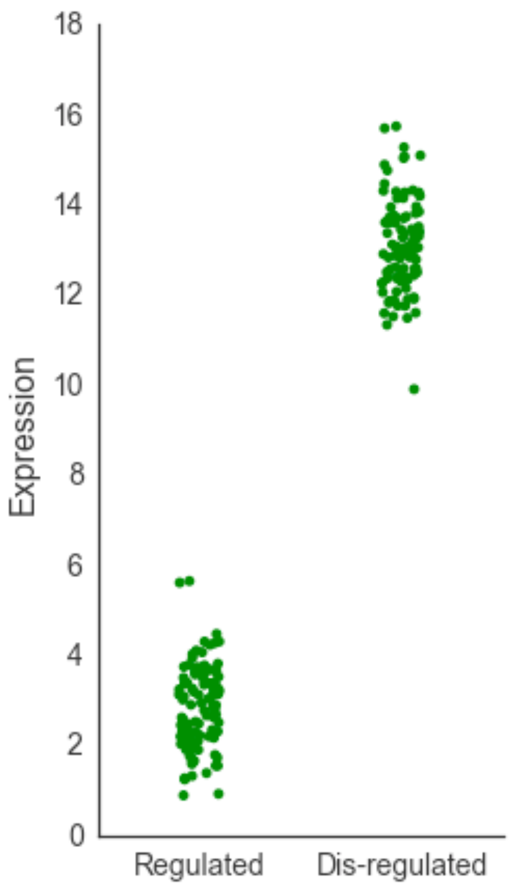
Joanna E. Long[a], Mark T. Drayson[b], Angela E. Taylor[d], Kai M. Toellner[b], Janet M. Lord[c, 1], Anna C. Phillips[a, 1], ,

Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes

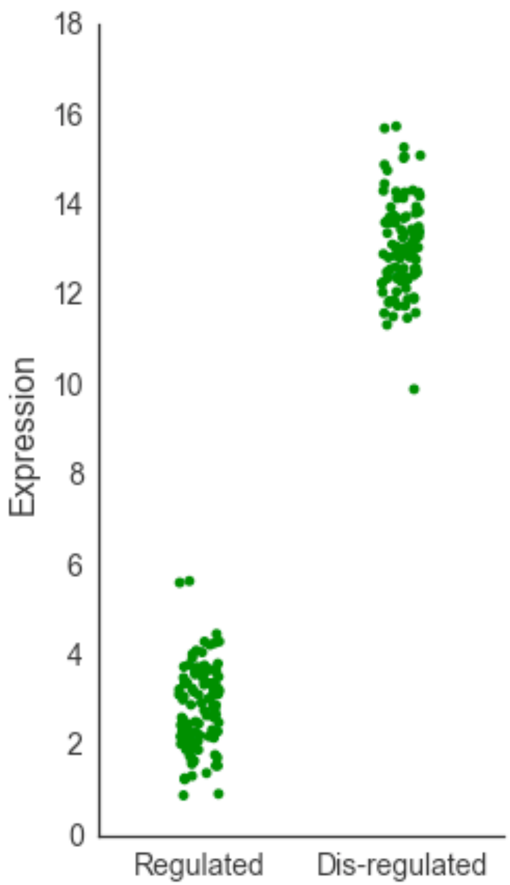
Biliana Marcheva Kathryn Moynihan Ramsey Ethan D. Buhr Yumiko Kobayashi Hong Su Caroline H. Ko Ganka Ivanova Chiaki Omura Shelley Mo Martha H. Vitaterna James P. Lopez Louis H. Philipson Christopher A. Bradfield Seth D. Crosby Lellean JeBailey Xiaozhong Wang Joseph S. Takahashi Joseph Bass

Dis-regulation of gene expression can cause physiological changes and disease

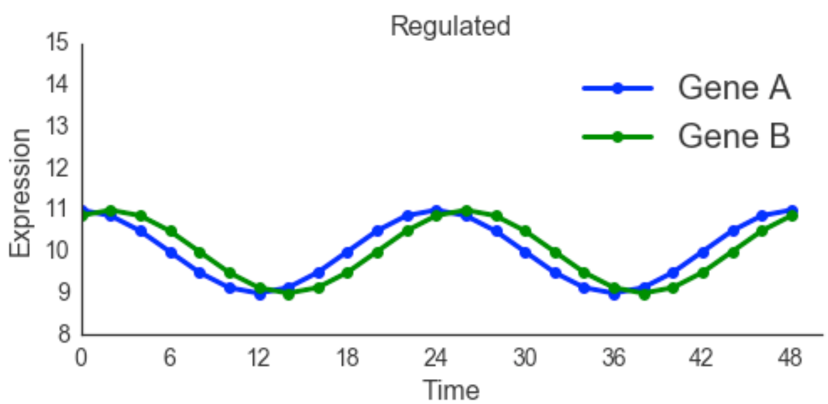
Dis-regulation can be due to changes in expression level



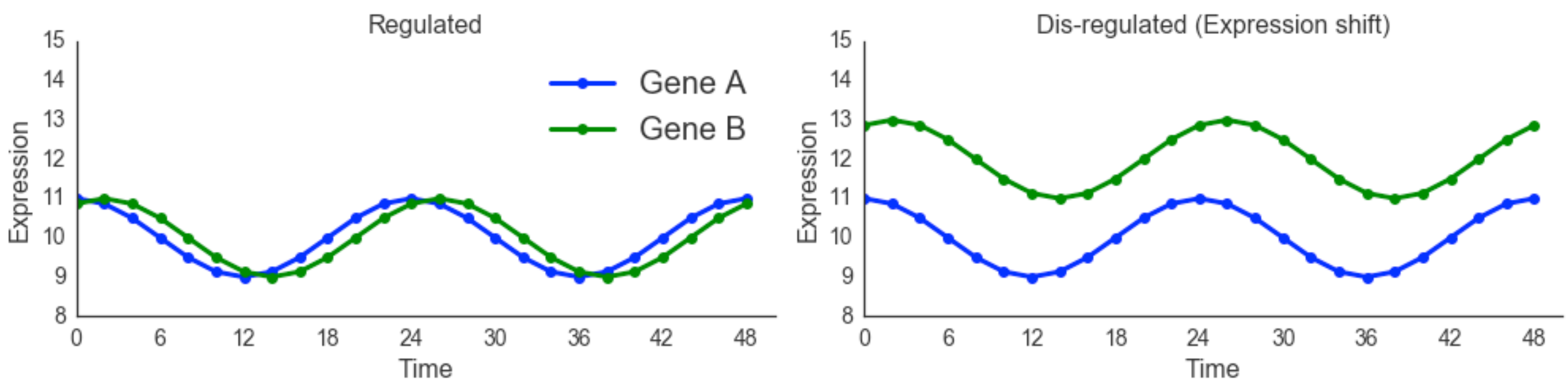
Dis-regulation can be more than changes in expression level



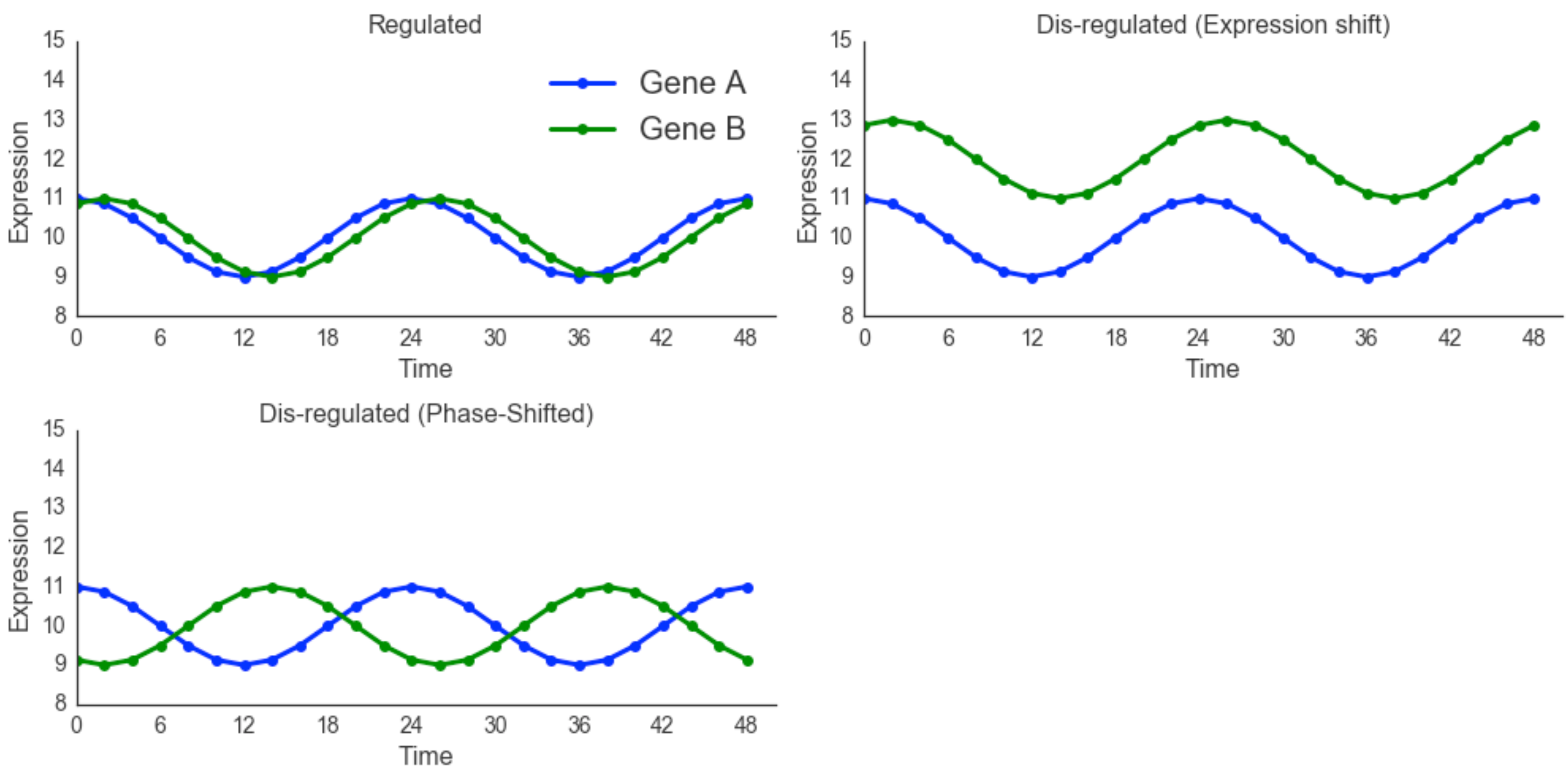
Dis-regulation can be more than changes in expression level



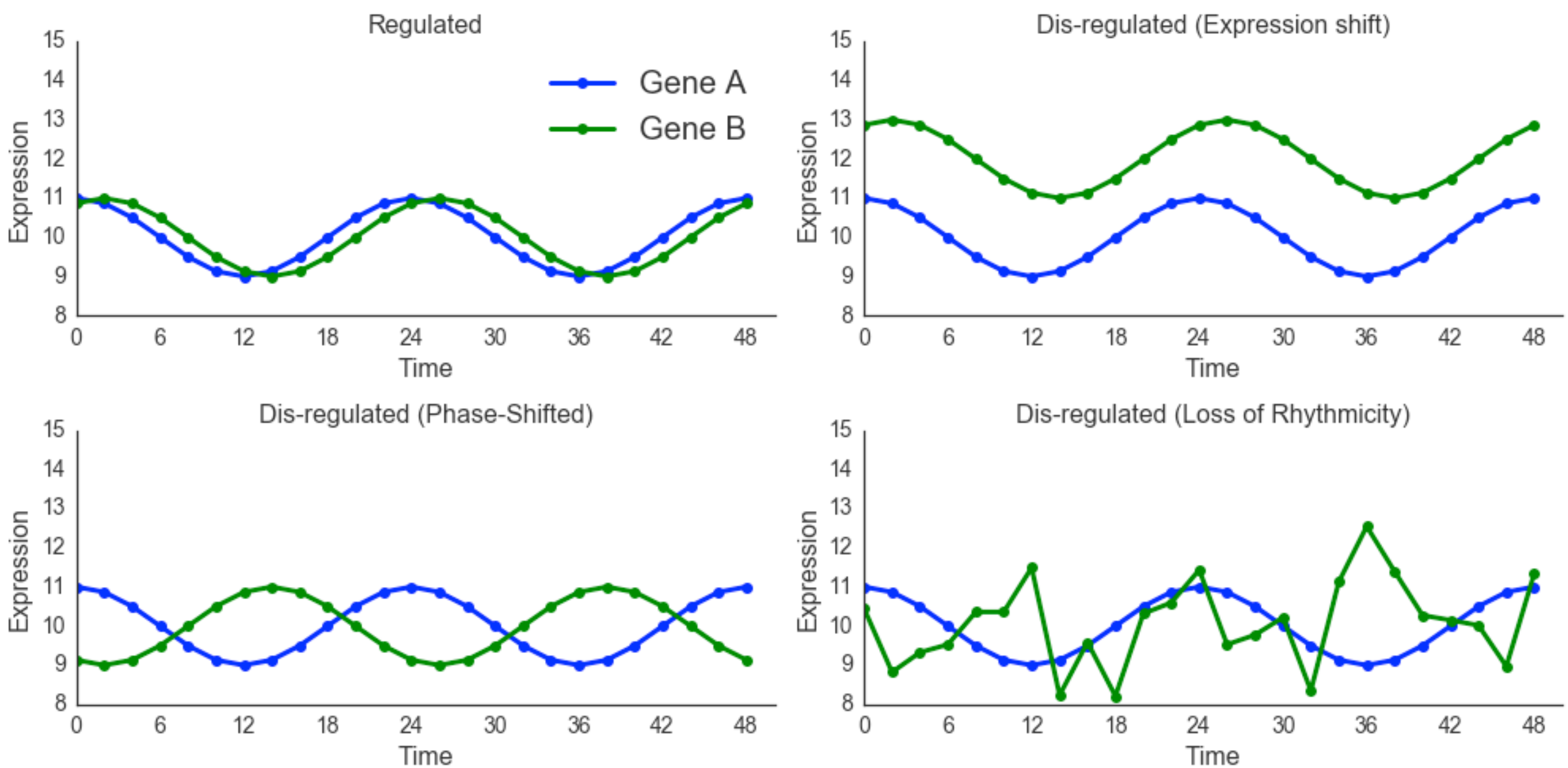
Dis-regulation can be more than changes in expression level



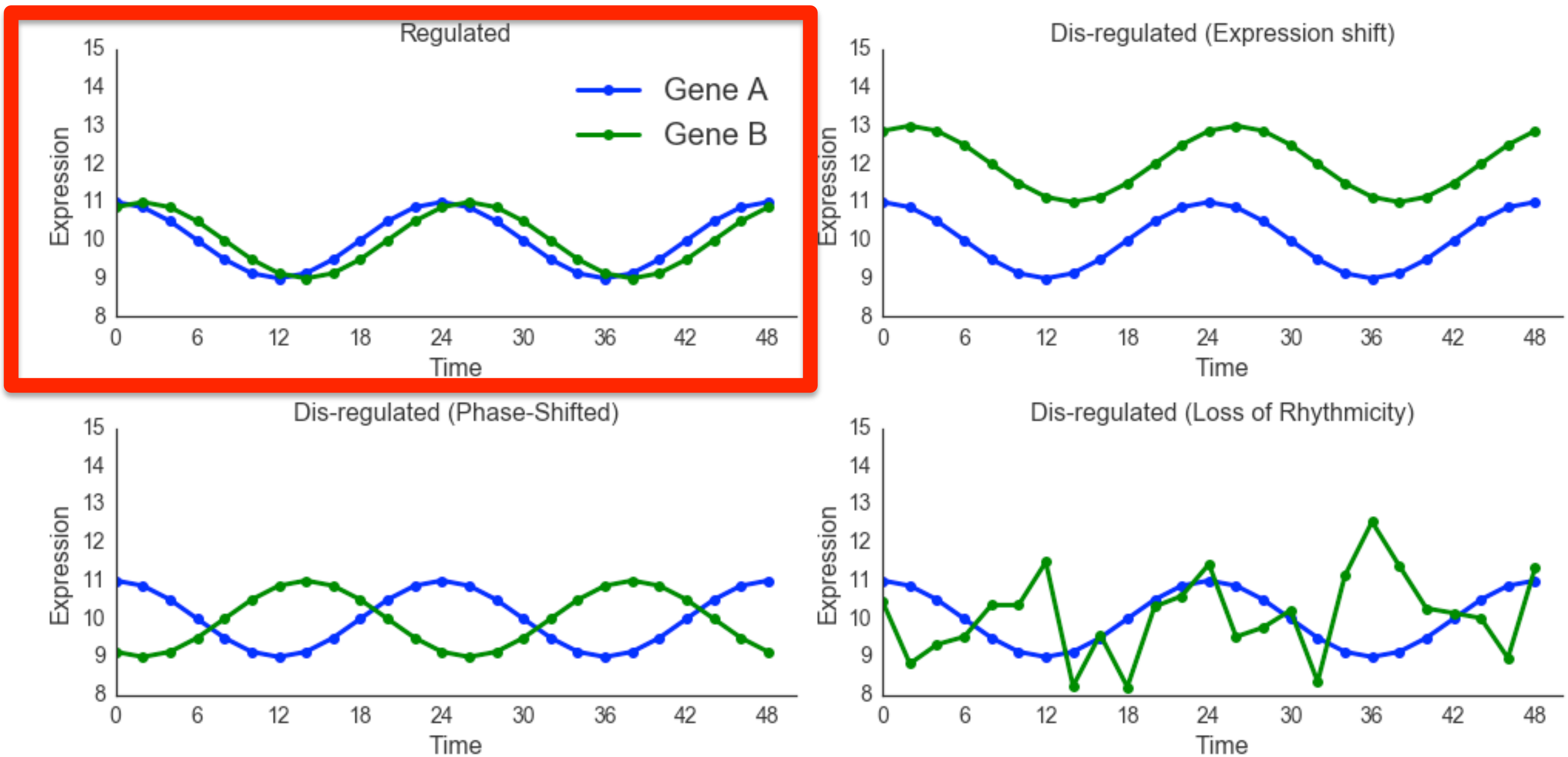
Dis-regulation can be more than changes in expression level



Dis-regulation can be more than changes in expression level



Dis-regulation can be more than changes in expression level



Outline

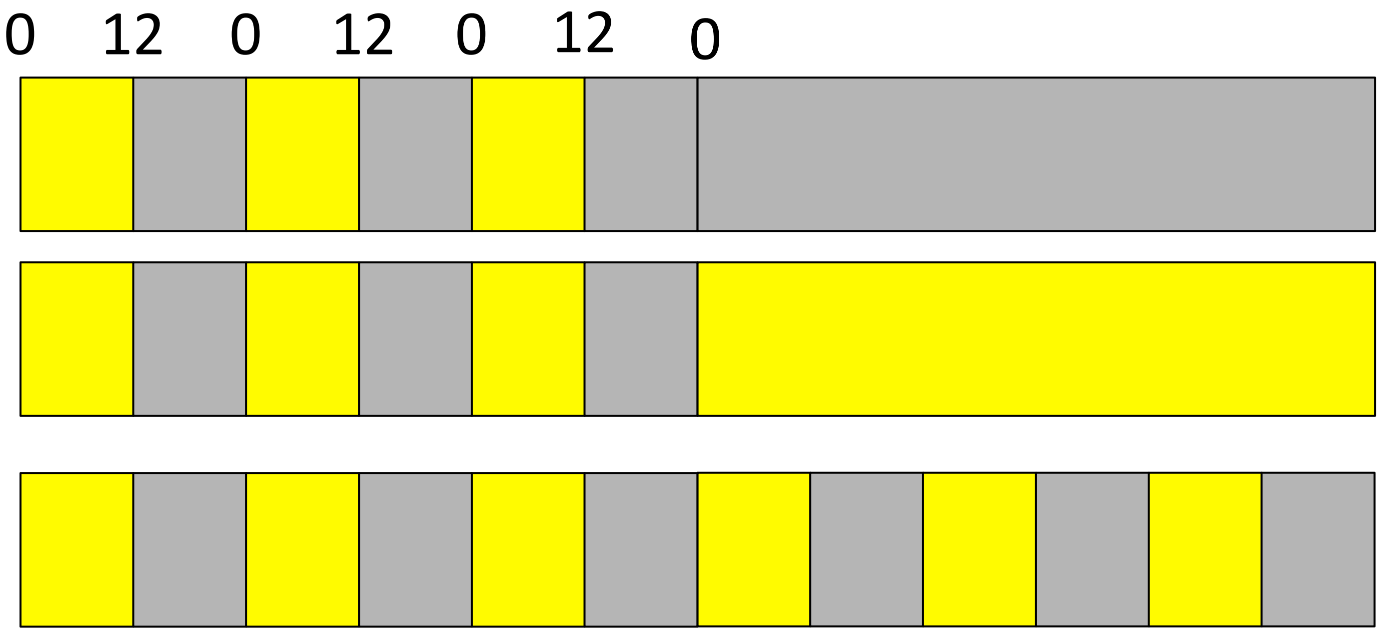
- Biological and Statistical Background
- Improvements to a Rhythm Detection method
- Comparing rhythmicity across conditions
- Future Directions

Outline

- Biological and Statistical Background
  - Circadian experiments
  - Challenges in rhythm detection
  - Current rhythm detection methods

Circadian experiment

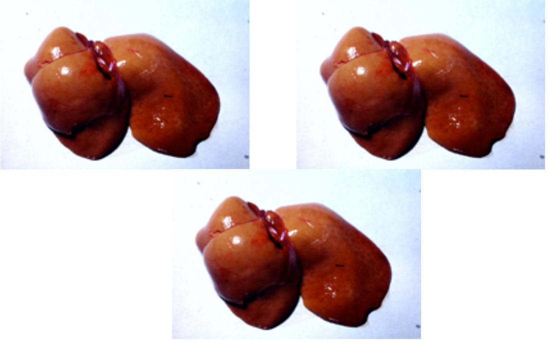
12 h light
12 h dark



Molecular circadian experiment

12 h light
12 h dark

http://animal-dream.com/data_images/mouse/mouse1.jpg



http://www.livercure.org/wp-content/uploads/2012/09/5b.jpg

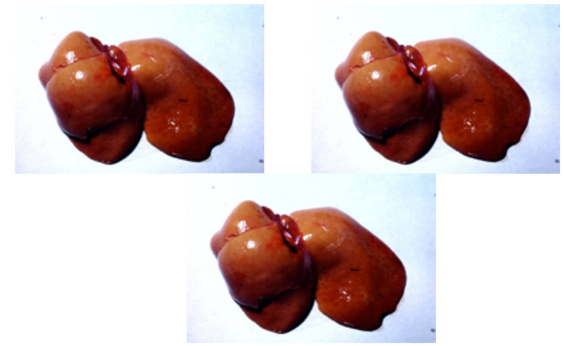
https://online-shop.eppendorf.com.my/MY-en/Laboratory-Consumables-44512/Tubes-44515/Eppendorf-Safe-Lock-Tubes-PF-8863.html

http://assets.illumina.com/content/dam/illumina-marketing/images/home/systems/web-graphics-hiseq2500-real.png

Molecular circadian experiment

12 h light
12 h dark

http://animal-dream.com/data_images/mouse/mouse1.jpg



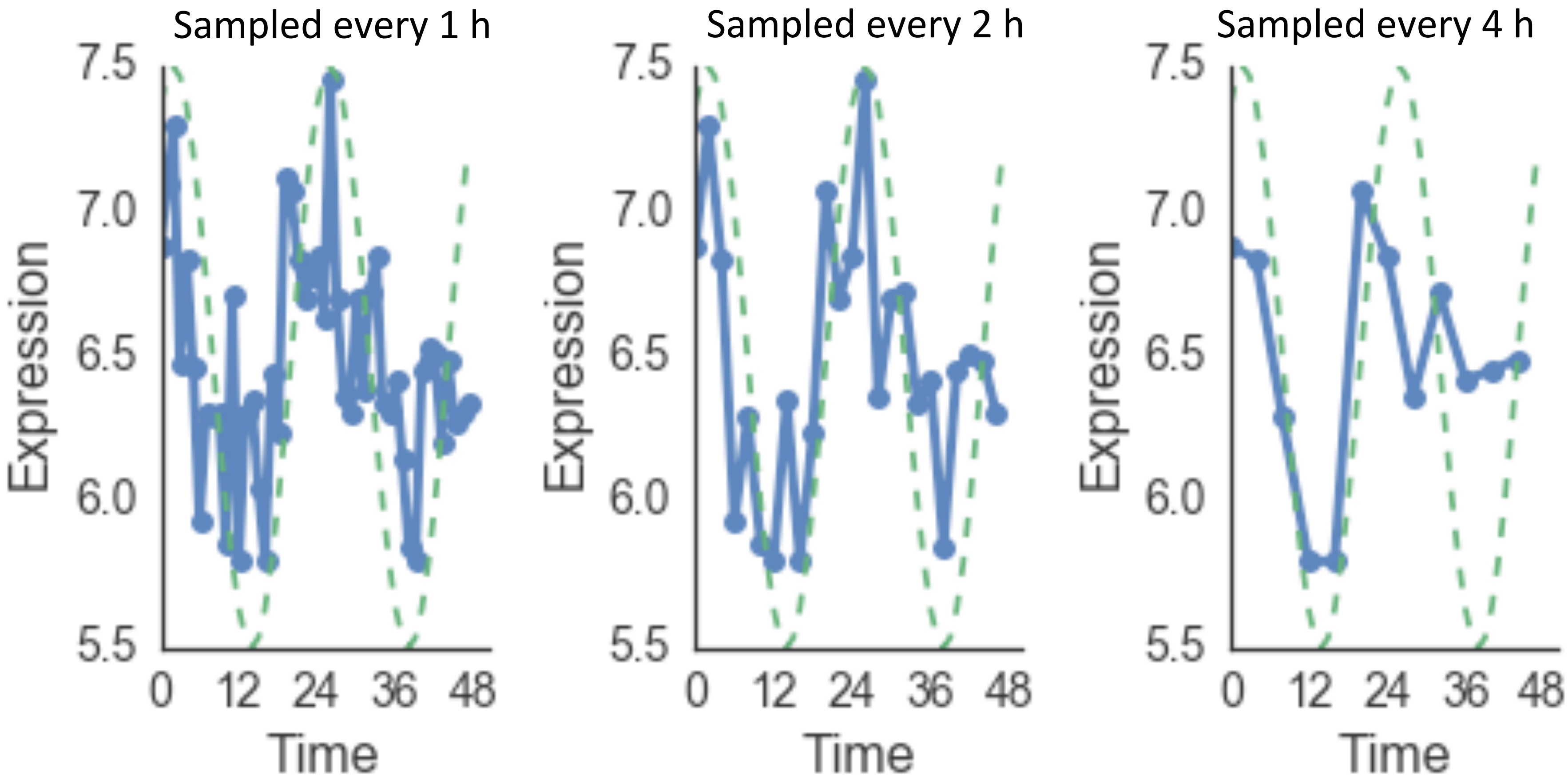
http://www.livercure.org/wp-content/uploads/2012/09/5b.jpg

per time point

https://online-shop.eppendorf.com.my/MY-en/Laboratory-Consumables-44512/Tubes-44515/Eppendorf-Safe-Lock-Tubes-PF-8863.html

http://assets.illumina.com/content/dam/illumina-marketing/images/home/systems/web-graphics-hiseq2500-real.png

Three challenges of rhythm detection



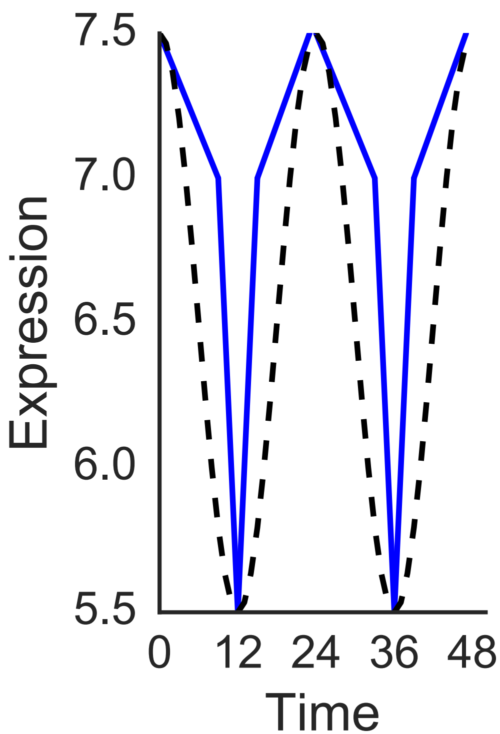
- Sparse sampling of data
- High noise of measurements
- High rate of arrhythmic genes

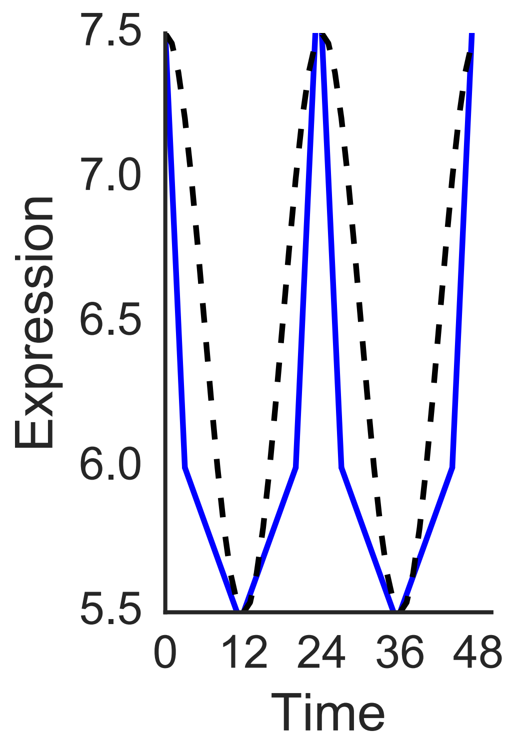
Time series data from Hughes *et al. PLoS Gen.* 2009

Rhythm detection approaches

- Cosine-fitting methods
  - COSOPT (Straume *et al.* 2004)
  - ARSER (Yang *et al.* 2010)
- Fourier-based methods
  - F24 (Wijnen *et al.* 2009)
- Reference-free methods
  - ANOVA (Keegan *et al.* 2007)
  - Cyclohedron test (Morton *et al.* 2007)
  - Address reduction (Fink *et al.* 2007)
  - Stable Persistence (Edelsbrunner *et al.* 2000)

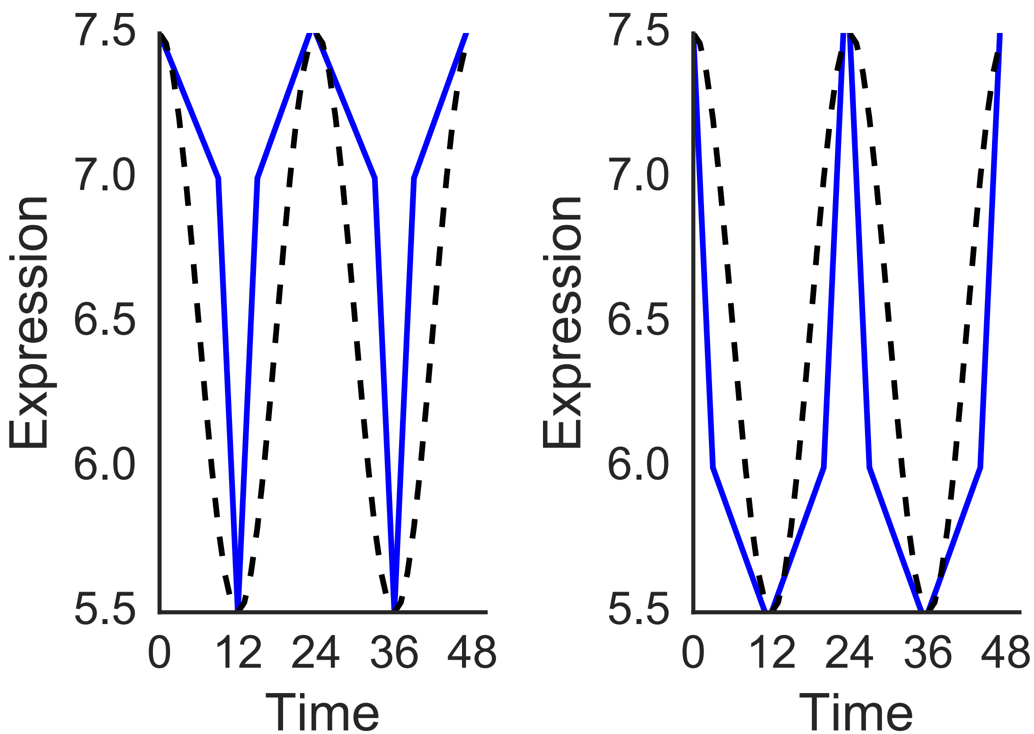
Rhythm detection approaches





- Cosine-fitting methods
  - COSOPT (Straume *et al.* 2004)
  - ARSER (Yang *et al.* 2010)
- Fourier-based methods
  - F24 (Wijnen *et al.* 2009)
- Reference-free methods
  - ANOVA (Keegan *et al.* 2007)
  - Cyclohedron test (Morton *et al.* 2007)
  - Address reduction (Fink *et al.* 2007)
  - Stable Persistence (Edelsbrunner *et al.* 2000)

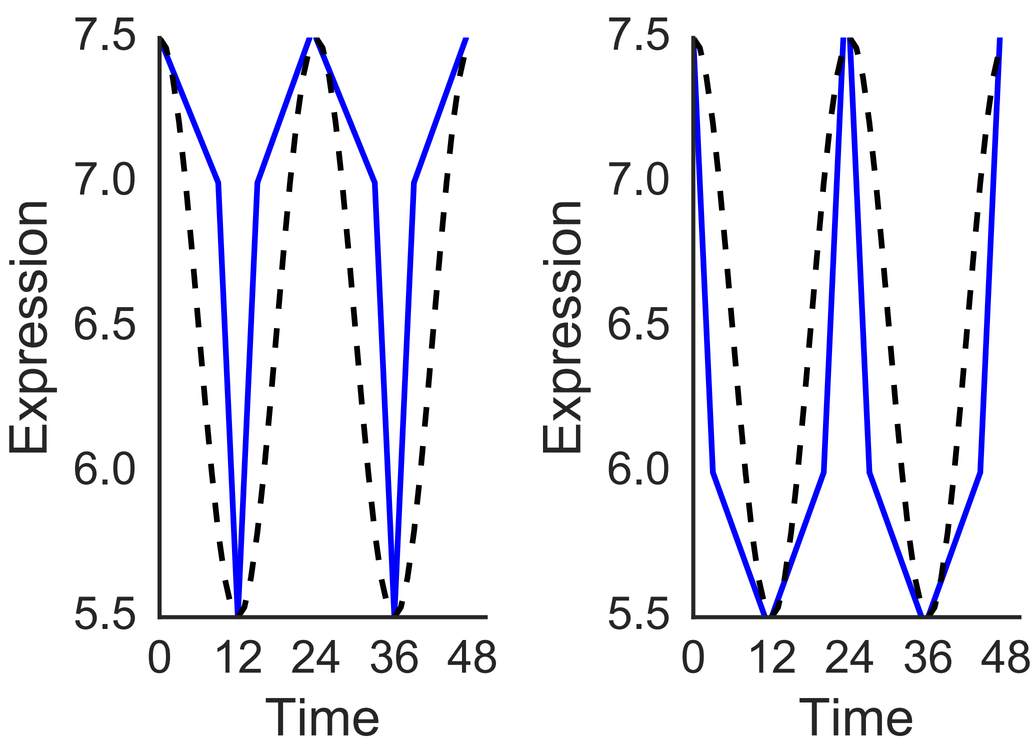
Rhythm detection approaches



- Non-parametric reference waveform methods
  - JTK_CYCLE (Hughes *et al.* 2010)
  - RAIN (Thaben *et al.* 2014)

- Cosine-fitting methods
  - COSOPT (Straume *et al.* 2004)
  - ARSER (Yang *et al.* 2010)
- Fourier-based methods
  - F24 (Wijnen *et al.* 2009)
- Reference-free methods
  - ANOVA (Keegan *et al.* 2007)
  - Cyclohedron test (Morton *et al.* 2007)
  - Address reduction (Fink *et al.* 2007)
  - Stable Persistence (Edelsbrunner *et al.* 2000)

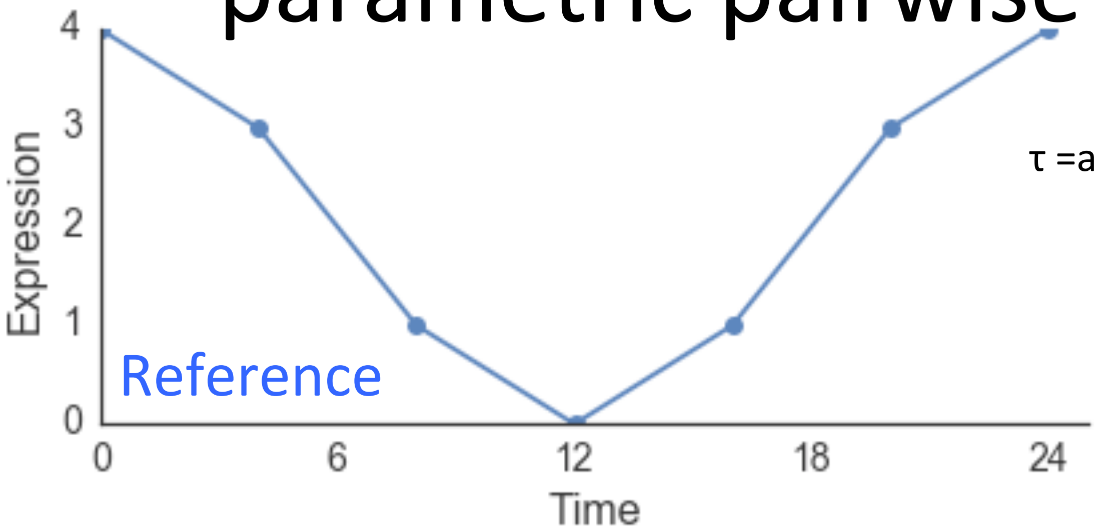
Rhythm detection approaches



- Non-parametric reference waveform methods
  - **JTK_CYCLE (Hughes *et al.* 2010)**
  - RAIN (Thaben *et al.* 2014)

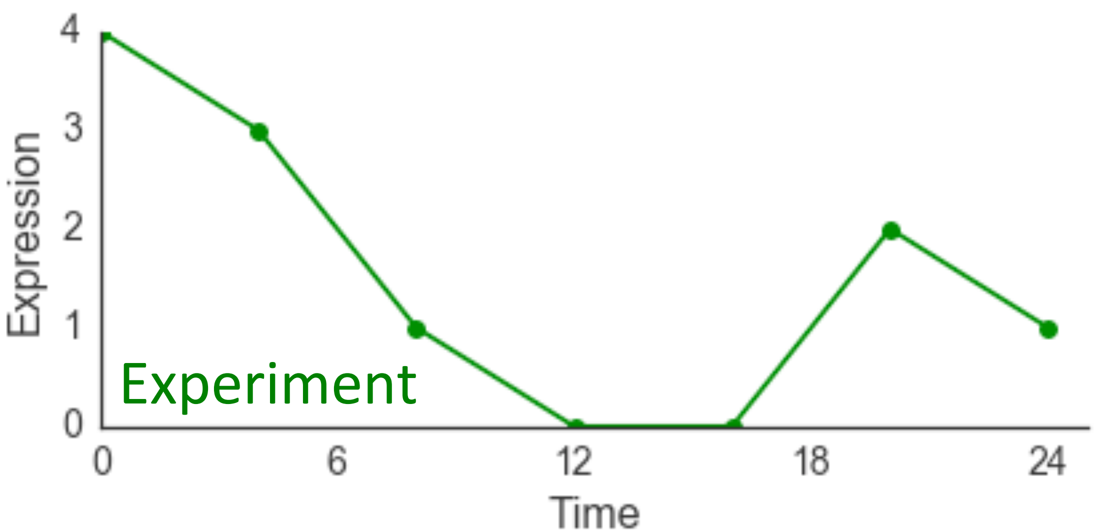
- Cosine-fitting methods
  - COSOPT (Straume *et al.* 2004)
  - ARSER (Yang *et al.* 2010)
- Fourier-based methods
  - F24 (Wijnen *et al.* 2009)
- Reference-free methods
  - ANOVA (Keegan *et al.* 2007)
  - Cyclohedron test (Morton *et al.* 2007)
  - Address reduction (Fink *et al.* 2007)
  - Stable Persistence (Edelsbrunner *et al.* 2000)

JTK_CYCLE uses Kendall's non-parametric pairwise correlation

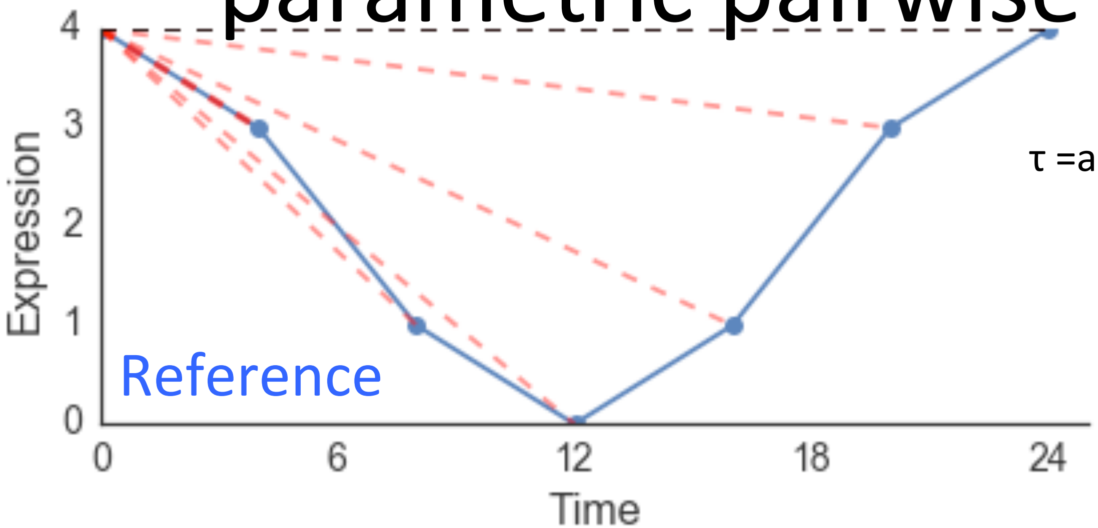


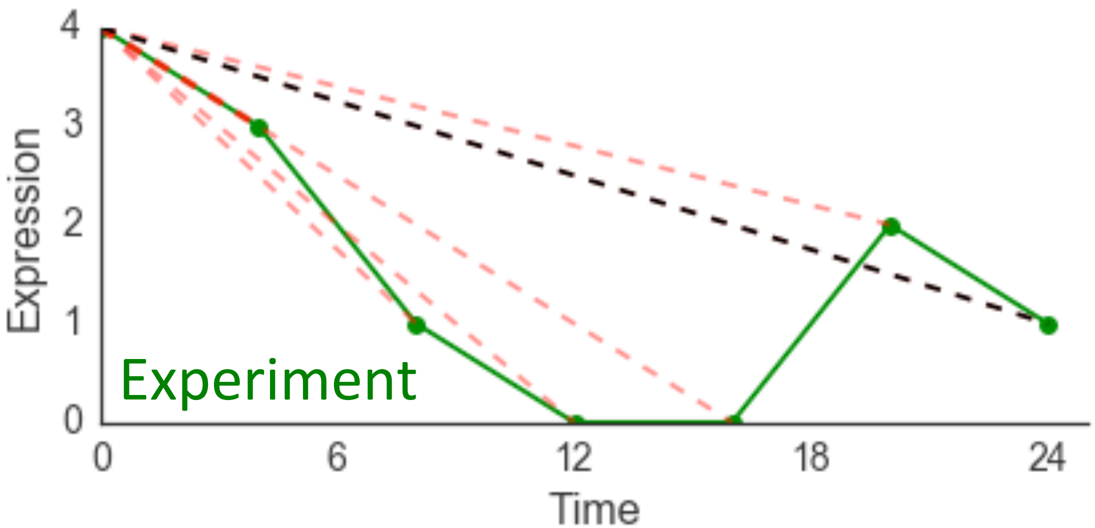
$$\tau = \text{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0			
1			
2			
3			
4			
5			
Sum			



JTK_CYCLE uses Kendall's non-parametric pairwise correlation

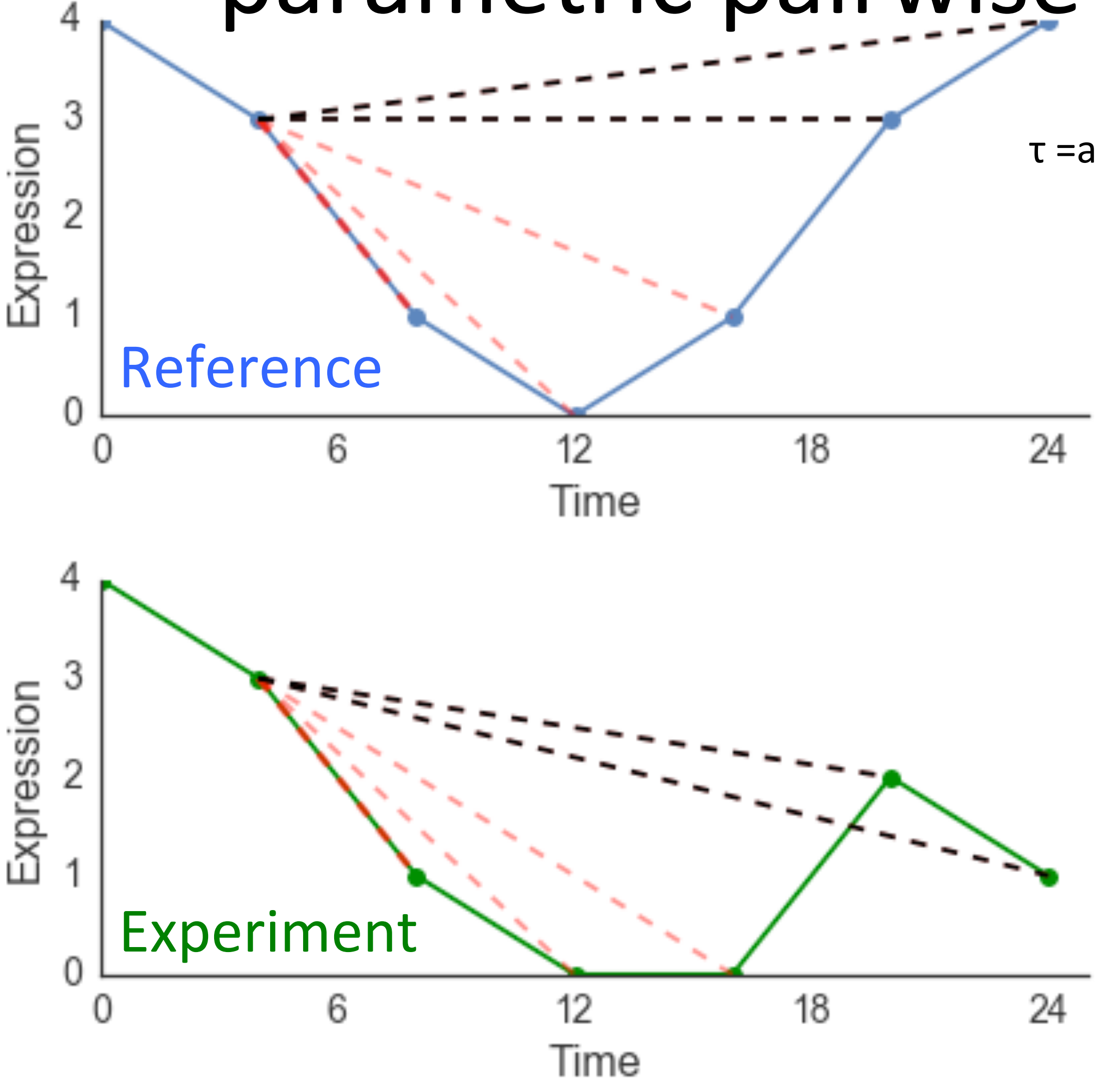




$$\tau = \operatorname{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1			
2			
3			
4			
5			
Sum			

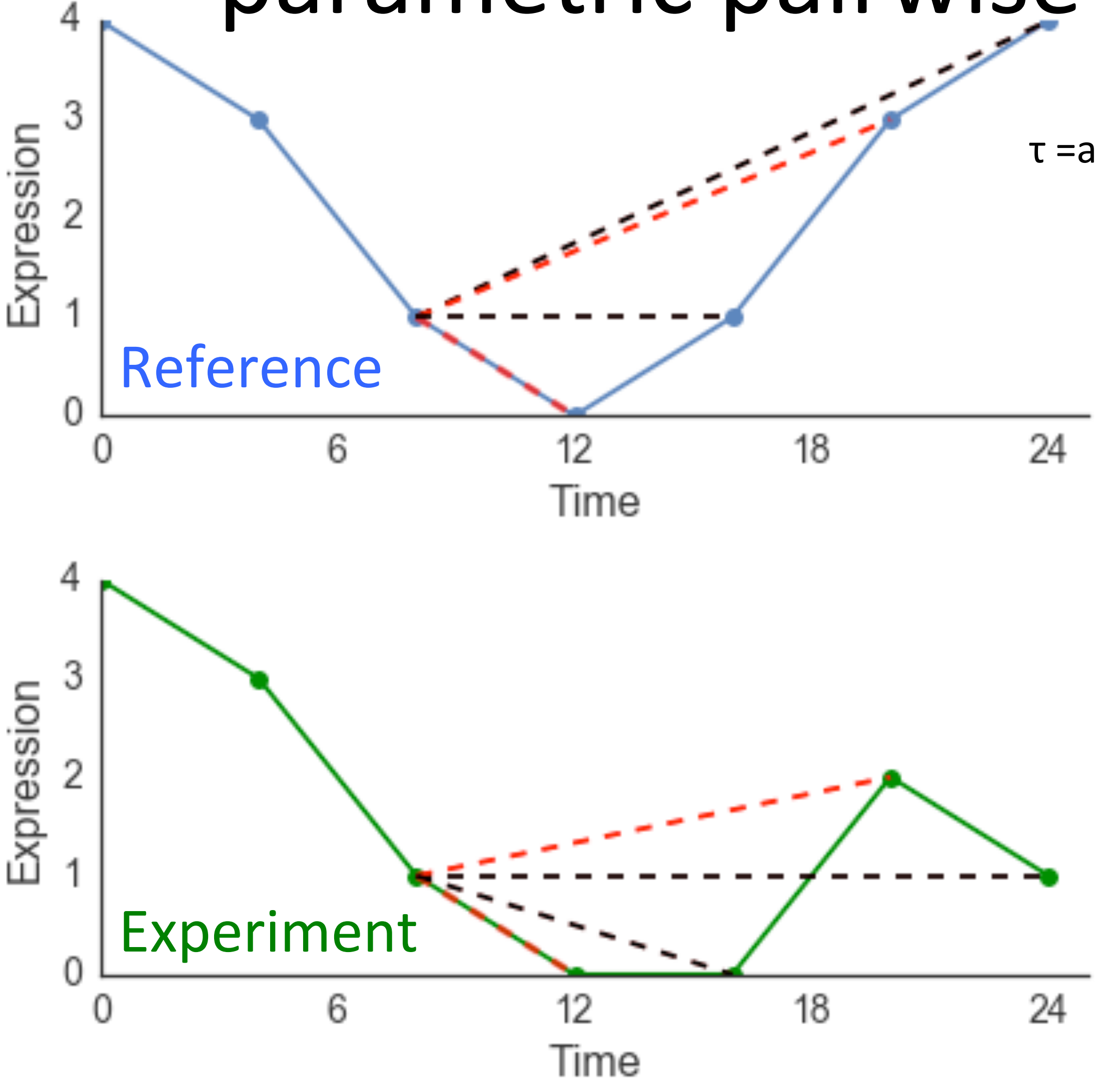
JTK_CYCLE uses Kendall's non-parametric pairwise correlation



$$\tau = \text{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2			
3			
4			
5			
Sum			

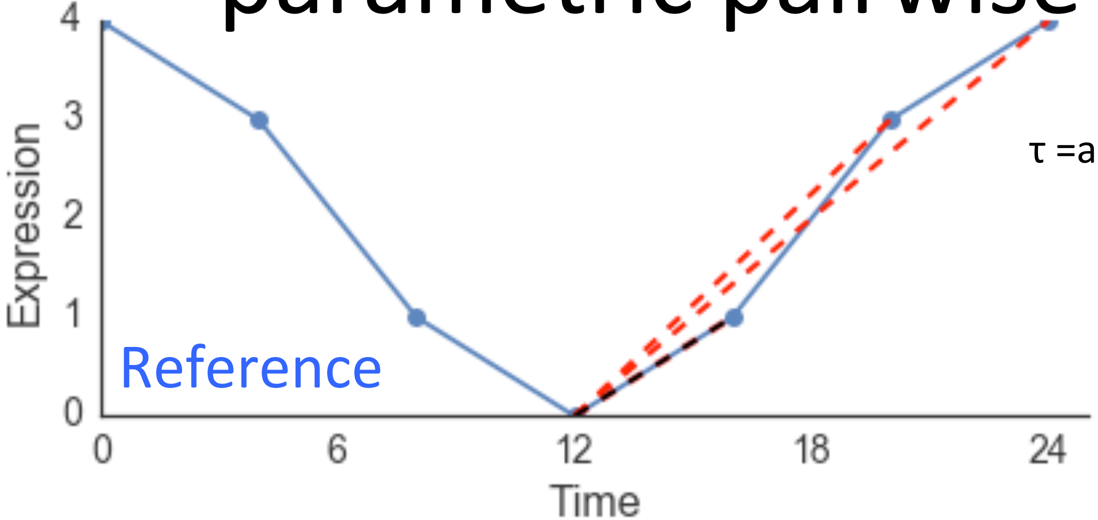
JTK_CYCLE uses Kendall's non-parametric pairwise correlation

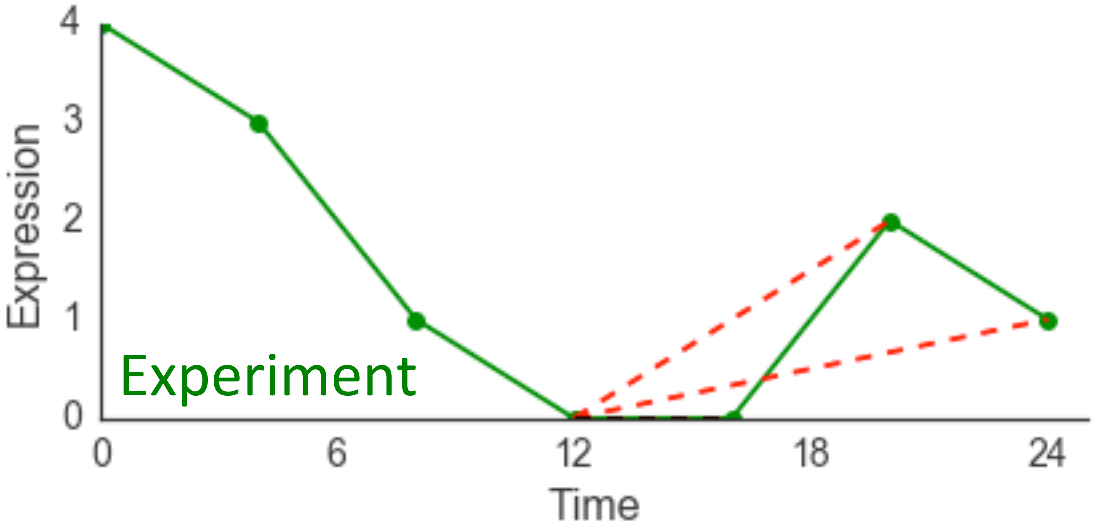


$$\tau = \operatorname{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3			
4			
5			
Sum			

JTK_CYCLE uses Kendall's non-parametric pairwise correlation

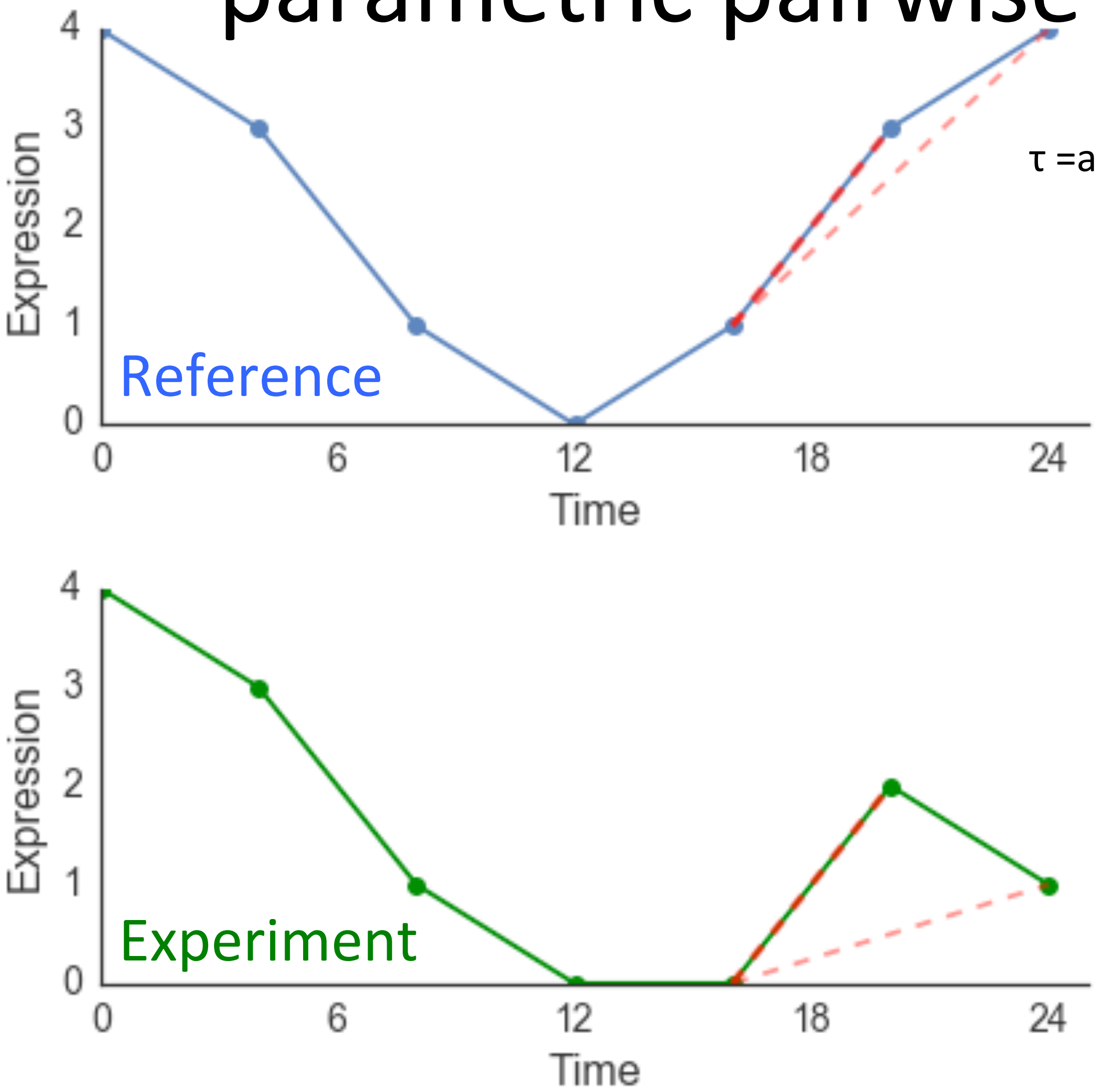




$$\tau = \operatorname{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3	2	0	1
4			
5			
Sum			

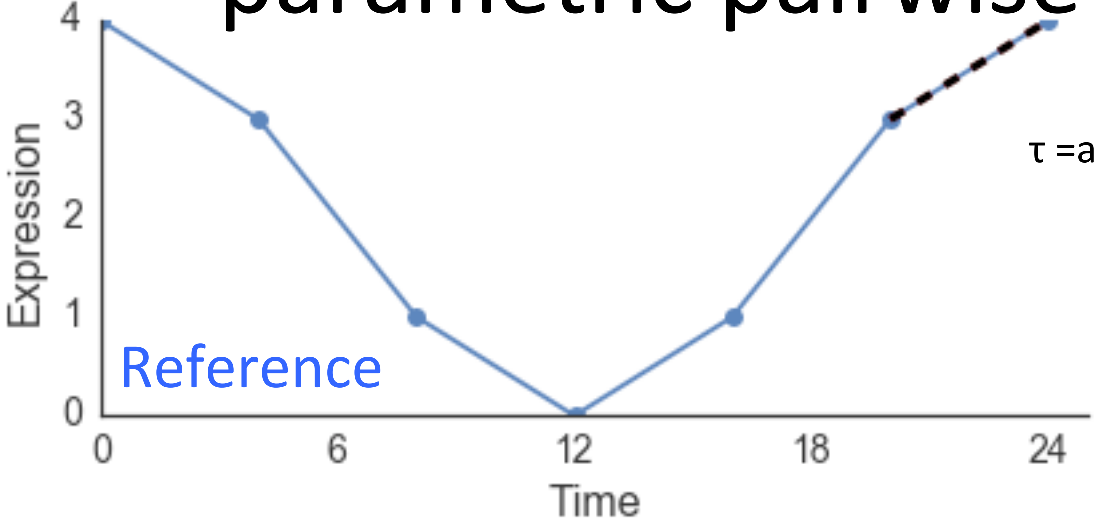
JTK_CYCLE uses Kendall's non-parametric pairwise correlation

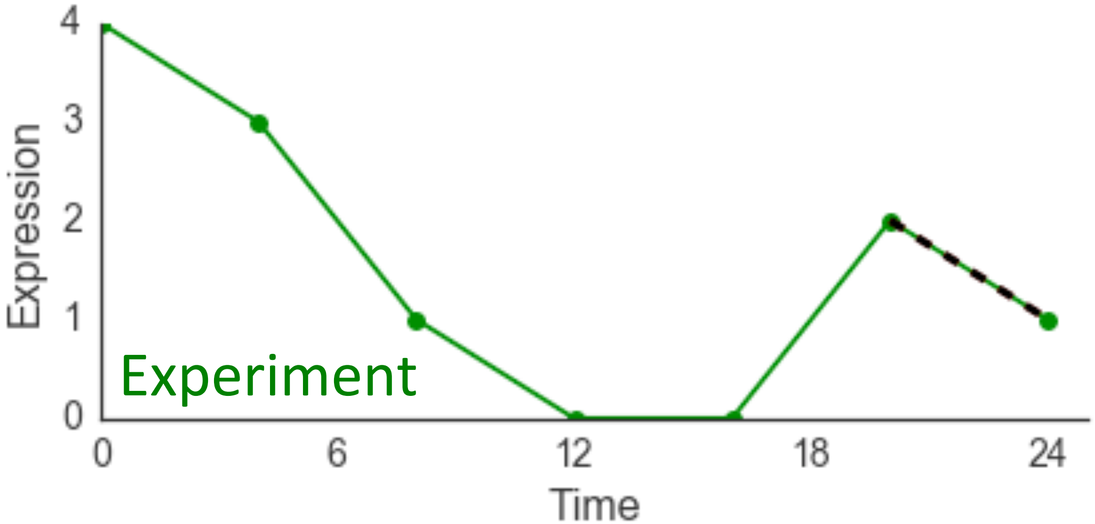


$$\tau = \operatorname{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3	2	0	1
4	2	0	0
5			
Sum			

JTK_CYCLE uses Kendall's non-parametric pairwise correlation

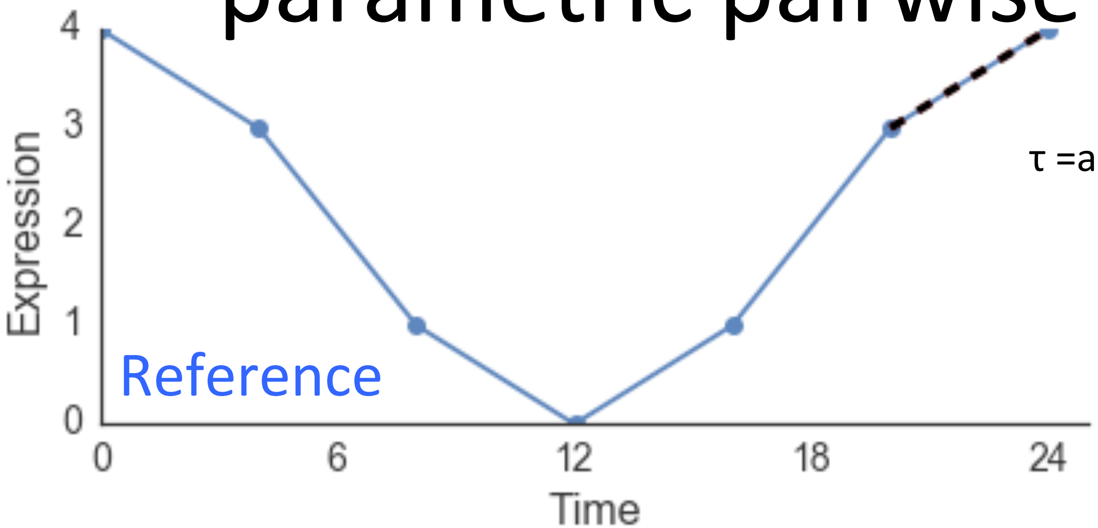


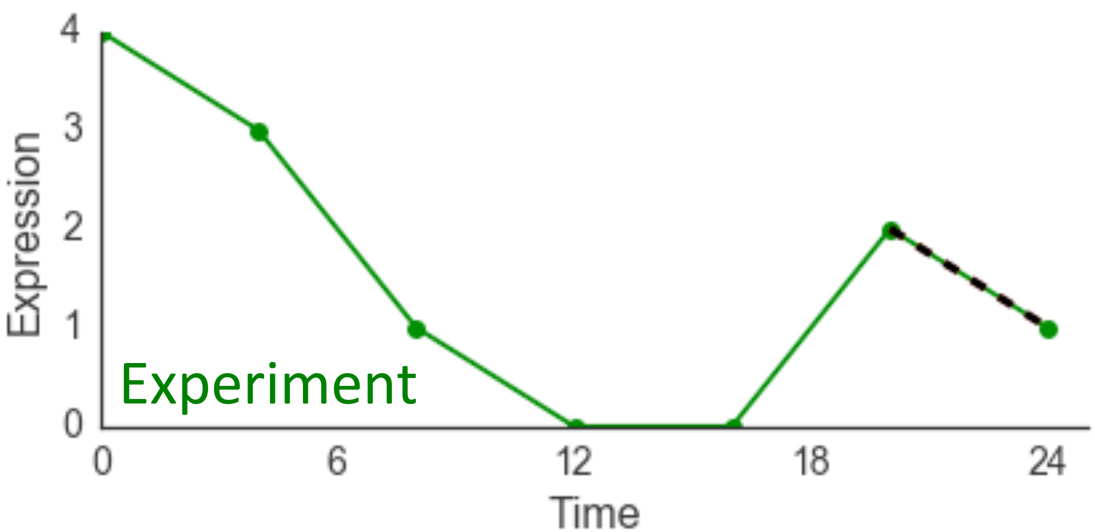


$$\tau = \operatorname{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3	2	0	1
4	2	0	0
5	0	1	0
Sum			

JTK_CYCLE uses Kendall's non-parametric pairwise correlation

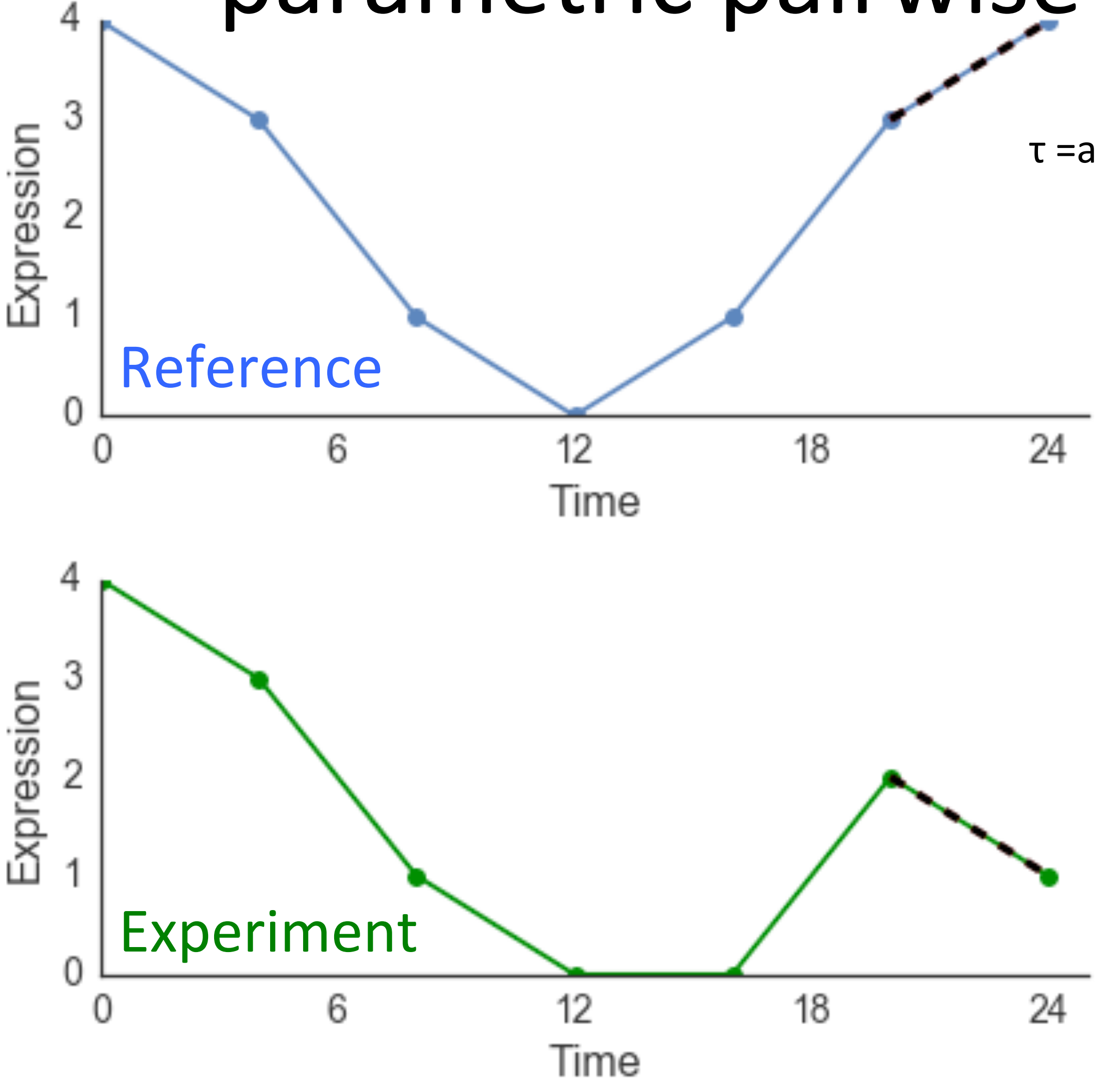




$$\tau = \text{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3	2	0	1
4	2	0	0
5	0	1	0
Sum	14	2	5

JTK_CYCLE uses Kendall's non-parametric pairwise correlation

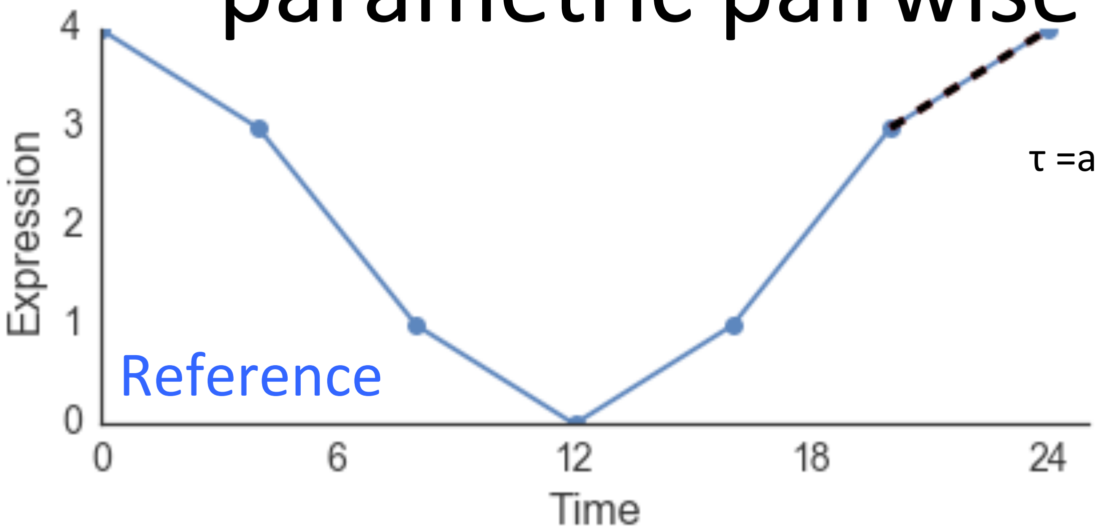


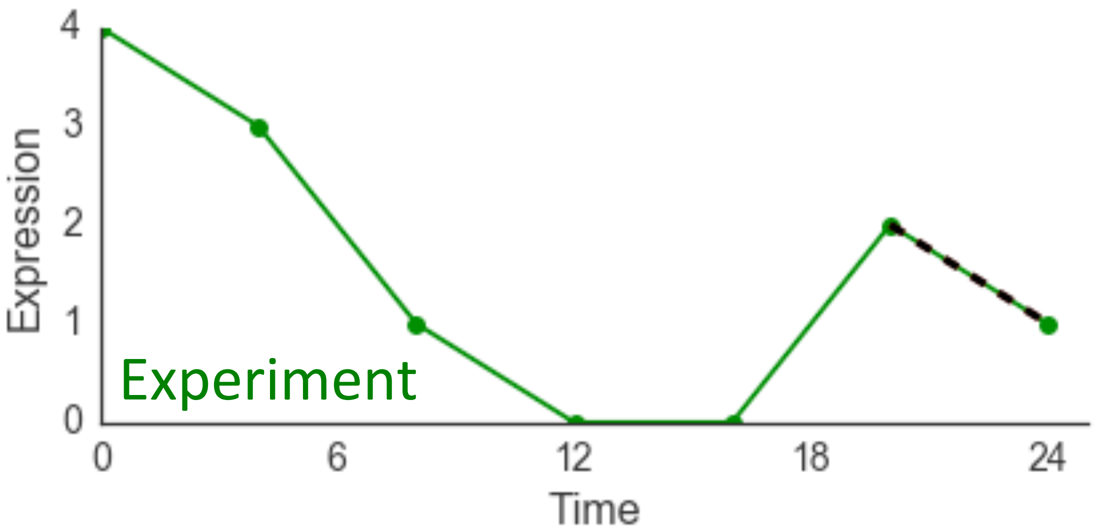
$$\tau = \text{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3	2	0	1
4	2	0	0
5	0	1	0
Sum	14	2	5

$\tau = \text{arctanh}((14\text{-}2)/21)$

JTK_CYCLE uses Kendall's non-parametric pairwise correlation





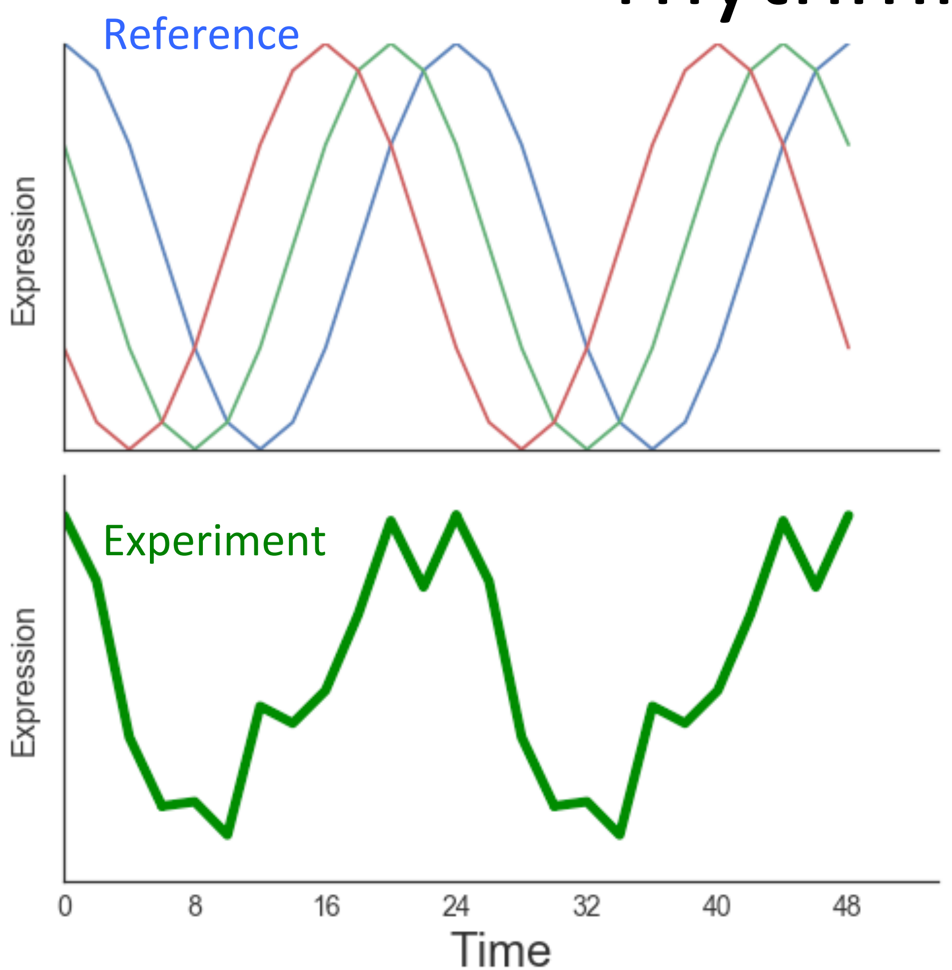
$$\tau = \operatorname{arctanh}\left(\frac{\text{Concordant pairs} - \text{Discordant pairs}}{\text{\# of pairs}}\right)$$

	Con.	Dis.	Zero
0	5	0	1
1	3	1	1
2	2	0	2
3	2	0	1
4	2	0	0
5	0	1	0
Sum	14	2	5

$\tau = \operatorname{arctanh}((14-2)/21)$

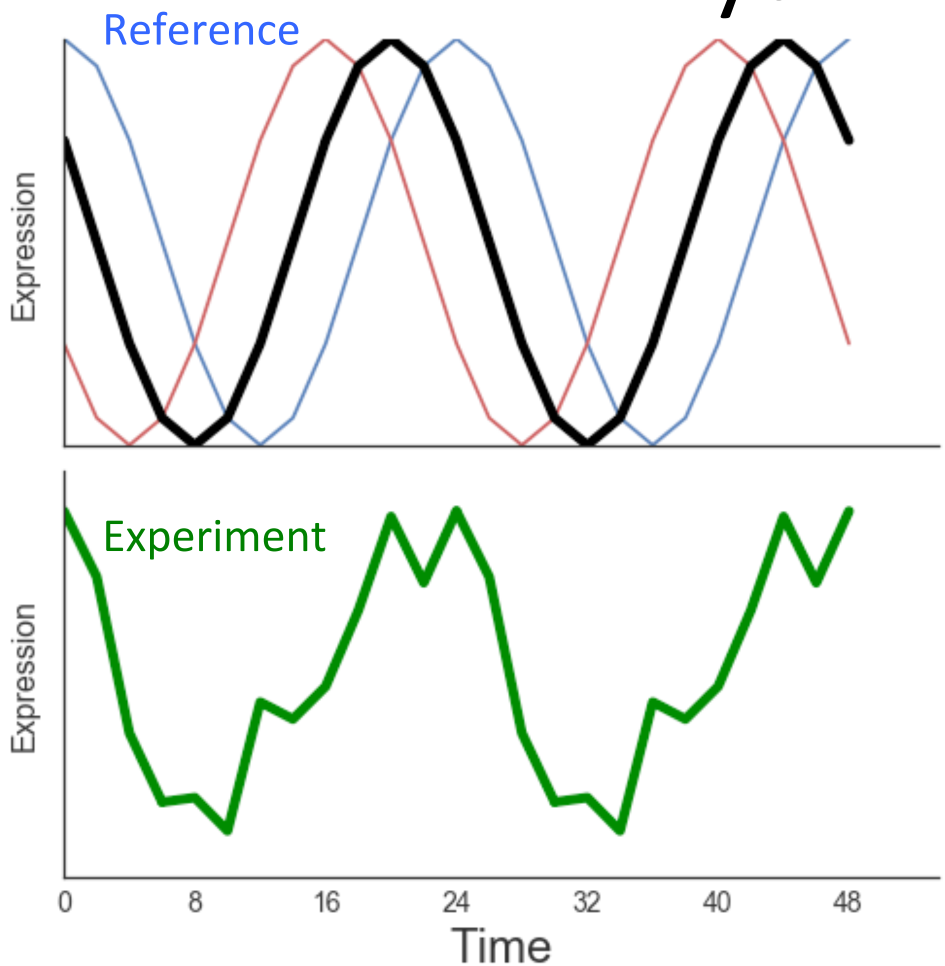
$\tau = 0.65$

JTK_CYCLE picks the best reference waveform match as its measure of rhythmicity



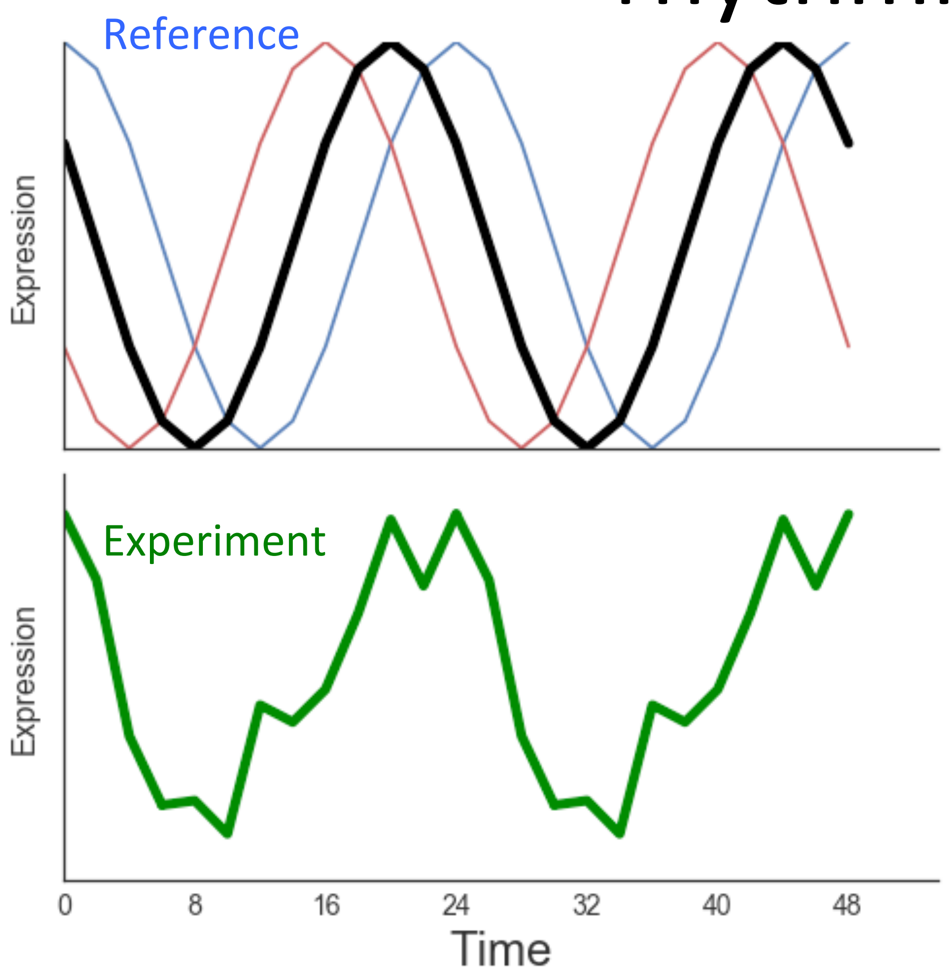
Period	Phase	τ
24	4	0.4
24	8	1.1
24	12	0.8

JTK_CYCLE picks the best reference waveform match as its measure of rhythmicity



Period	Phase	τ
24	4	0.4
24	**8**	**1.1**
24	12	0.8

JTK_CYCLE picks the best reference waveform match as its measure of rhythmicity



Period	Phase	τ
24	4	0.4
24	**8**	**1.1**
24	12	0.8

Data sampled every 2 h over 24 h:
12 possible phases

Outline

- Biological and Statistical Background
  - Circadian experiments
  - Challenges in rhythm detection
  - Current rhythm detection methods

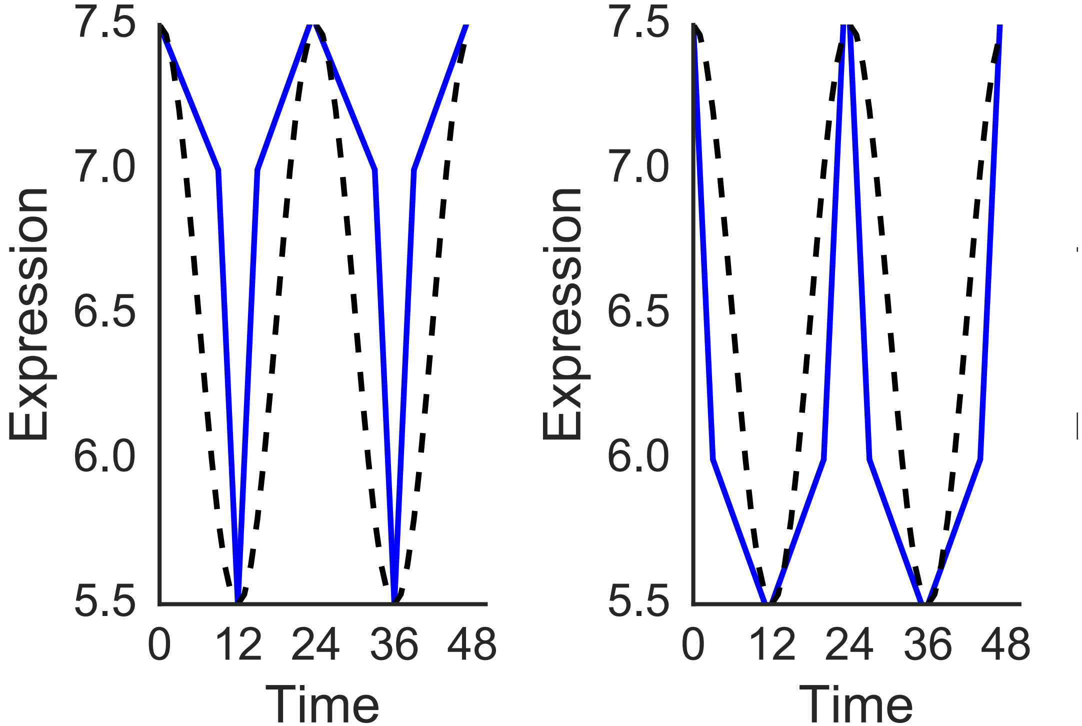
Outline

- Biological and Statistical Background
- **Improvements to JTK_CYCLE**
  - **Empirical JTK_CYCLE (eJTK)**
  - **Bootstrap eJTK (BooteJTK)**

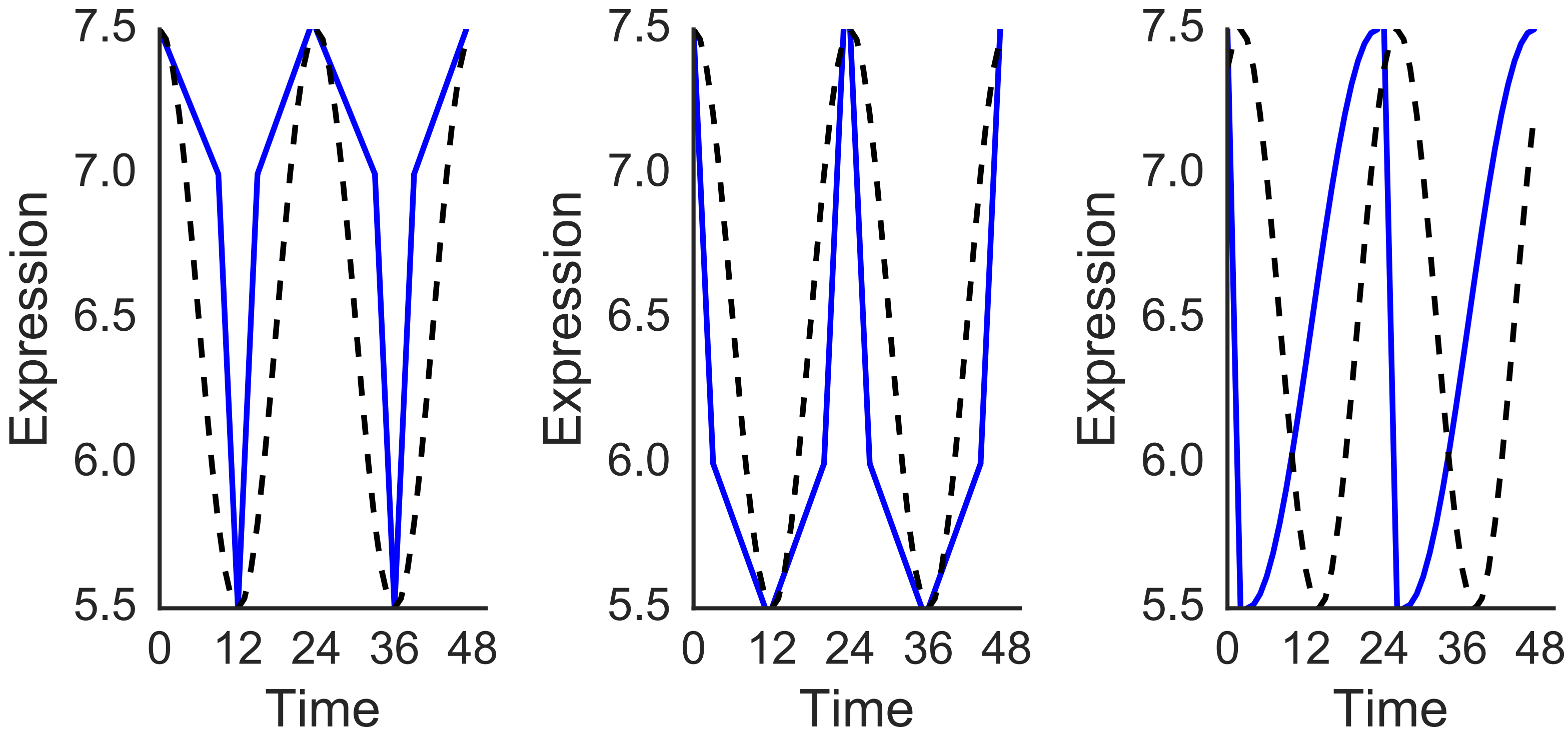
Outline

- Biological and Statistical Background
- Improvements to JTK_CYCLE
  - **Empirical JTK_CYCLE (eJTK)**
    - **Searching for asymmetric waveforms**
    - **Calculating accurate p-values**

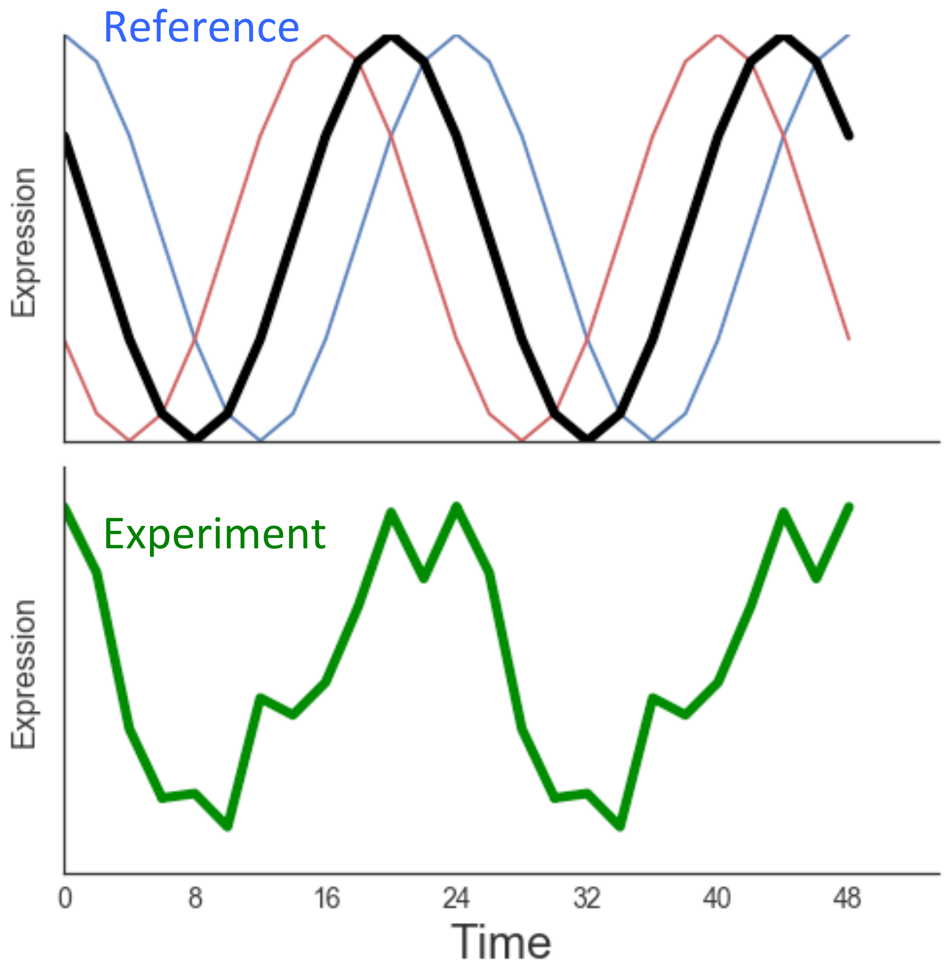
Traditionally, JTK_CYCLE did not search for asymmetric waveforms



Traditionally, JTK_CYCLE did not search for asymmetric waveforms

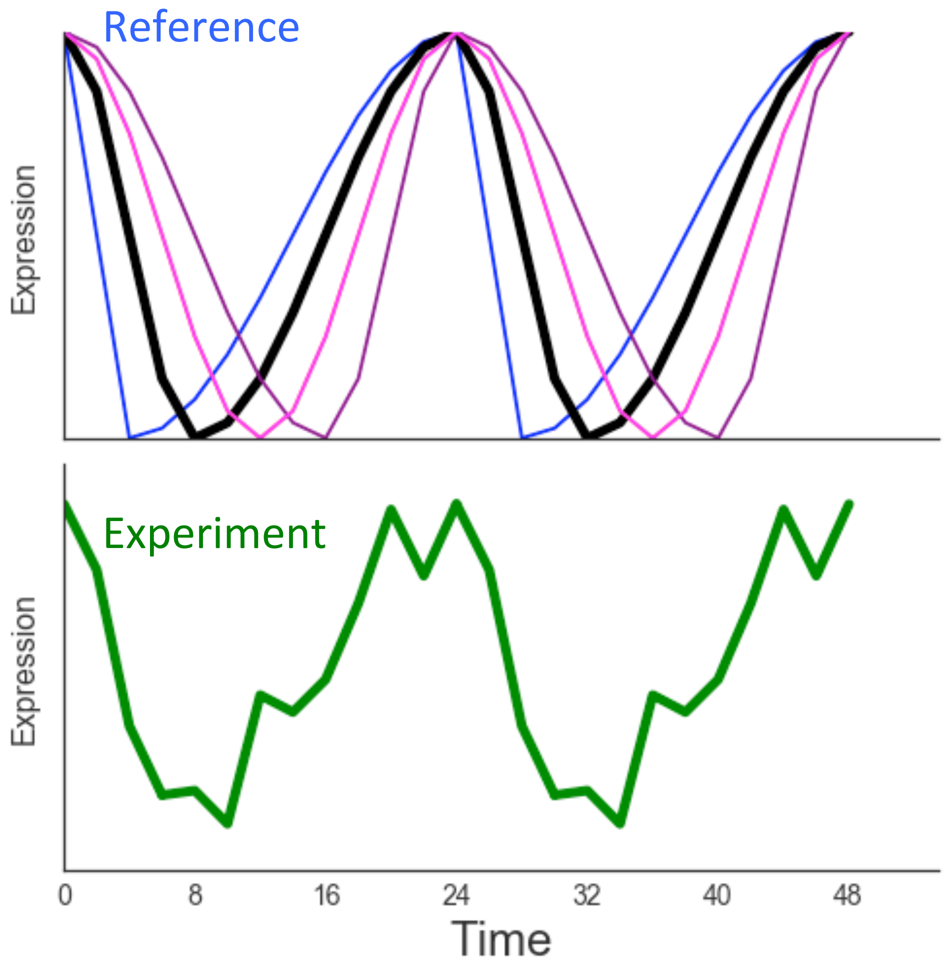


Asymmetric waveforms improve rhythm detection sensitivity



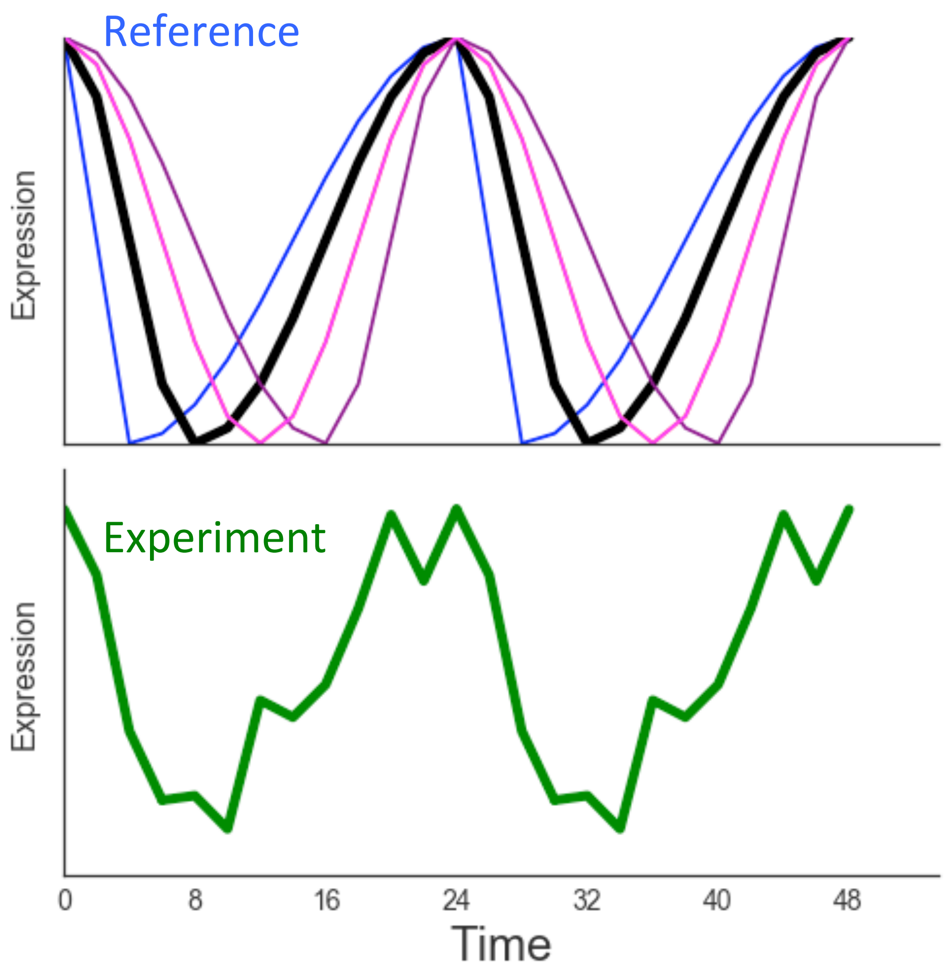
Period	Phase	τ
24	8	0.4
24	**10**	**1.1**
24	12	0.8

Asymmetric waveforms improve rhythm detection sensitivity



Period	Phase	Asym.	τ
24	10	4	0.6
24	**10**	**8**	**1.3**
24	10	12	1.0
24	10	16	0.9

Asymmetric waveforms improve rhythm detection sensitivity



Period	Phase	Asym.	τ
24	10	4	0.6
24	**10**	**8**	**1.3**
24	10	12	1.0
24	10	16	0.9

Data sampled every 2 h over 24 h:

12 possible phases

11 possible asymmetries

132 reference waveforms

P-values

P-values

- We do not know if what we observe is signal or just noise.

P-values

- We do not know if what we observe is signal or just noise.
- What is the probability that we see what we do if it were just noise?

P-values

- We do not know if what we observe is signal or just noise.
- What is the probability that we see what we do if it were just noise?
- If the probability is very low, then we intuit it is unlikely that it comes just from noise and is worth pursuing further.

P-values

- We do not know if what we observe is signal or just noise.
- What is the probability that we see what we do if it were just noise?
- If the probability is very low, then we intuit it is unlikely that it comes just from noise and is worth pursuing further.

Two ways of generating a p-value for a statistical summary of data

1. Analytical calculation
2. Via simulation

P-values

- We do not know if what we observe is signal or just noise.
- What is the probability that we see what we do if it were just noise?
- If the probability is very low, then we intuit it is unlikely that it comes just from noise and is worth pursuing further.

Two ways of generating a p-value for a statistical summary of data

1. Analytical calculation
2. **Via simulation**

American Statistical Association definition of p-values

P-values for JTK_CYCLE

P-values for JTK_CYCLE

"Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value."
- The ASA's Statement on p-values: Context, Process, and Purpose, 2016

P-values for JTK_CYCLE

The time series is generated from noise with no underlying signal

"Informally, a p-value is the **probability under a specified statistical model** that a statistical summary of the data would be equal to or more extreme than its observed value."
- The ASA's Statement on p-values: Context, Process, and Purpose, 2016

P-values for JTK_CYCLE

The time series is generated from noise
with no underlying signal

"Informally, a p-value is the **probability under a specified statistical model** that **a statistical summary of the data** would be equal to or more extreme than its observed value."
- The ASA's Statement on p-values: Context, Process, and Purpose, 2016

The Kendall's Tau correlation for several
reference waveforms where we then
pick the best correlation

P-values for JTK_CYCLE

The time series is generated from noise
with no underlying signal

"Informally, a p-value is the **probability under a specified statistical model** that **a statistical summary of the data** would **be equal to or more extreme than its observed value.**"
- The ASA's Statement on p-values: Context, Process, and Purpose, 2016

The Kendall's Tau correlation for several
reference waveforms where we then
pick the best correlation

Calculating p-values from simulated data

a statistical summary of the data

a specified statistical model

be equal to or more extreme than its observed value

Calculating p-values from simulated data

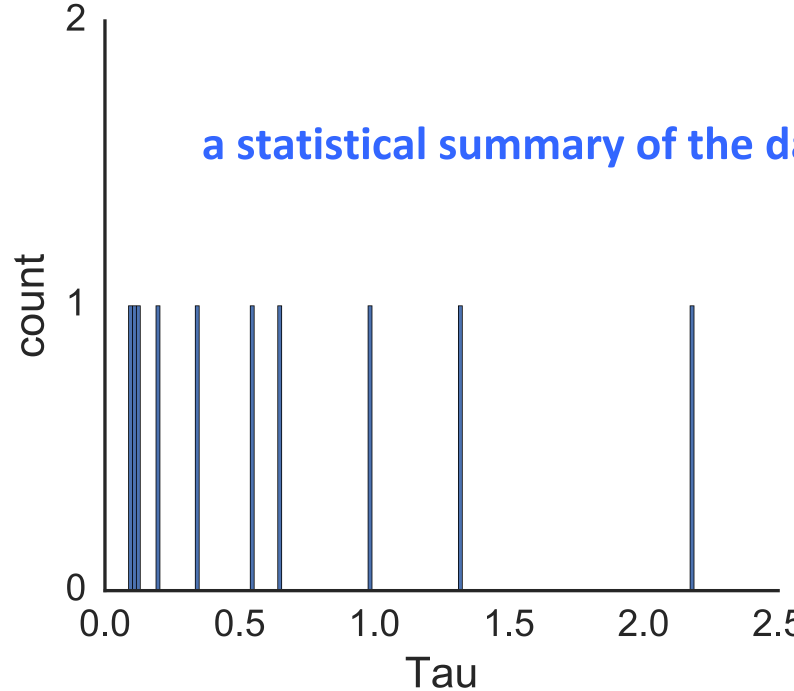
a statistical summary of the data

a specified statistical model

be equal to or more extreme than its observed value

Calculating p-values from simulated data

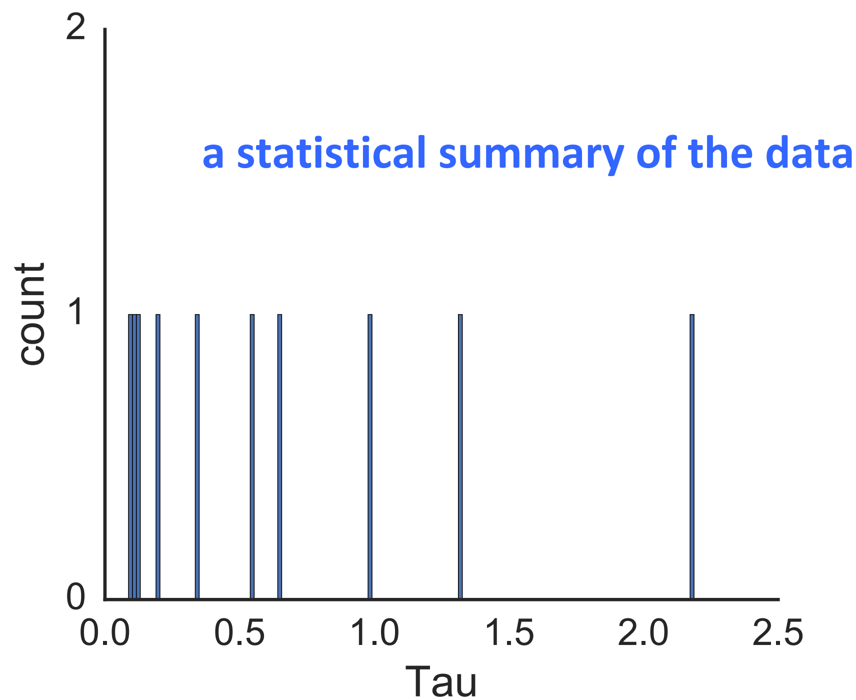
a specified statistical model



a statistical summary of the data

be equal to or more extreme than its observed value

Calculating p-values from simulated data

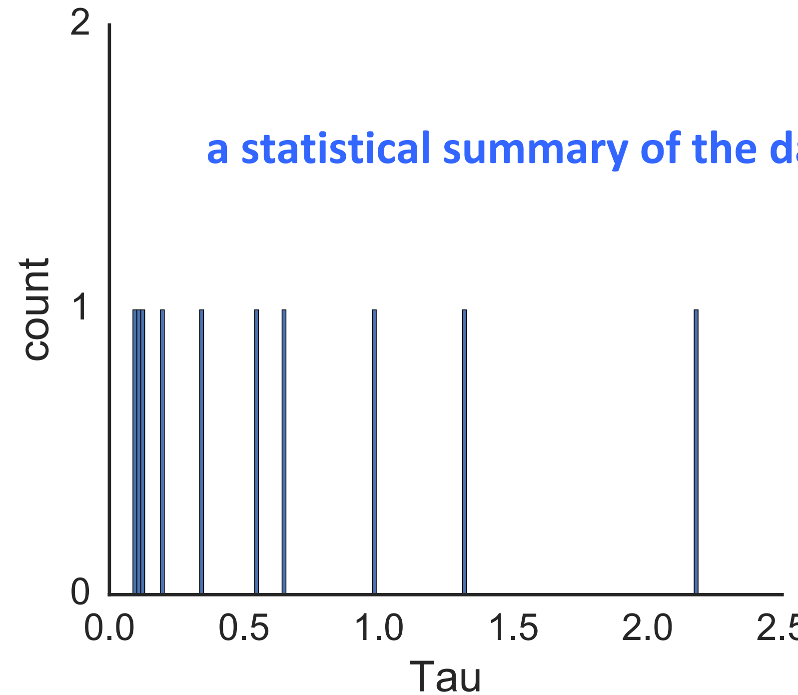


be equal to or more extreme than its observed value

$$\text{p-value} = \frac{(\# \geq \text{observed value})}{(\text{Total } \#)}$$

Calculating p-values from simulated data

a specified statistical model



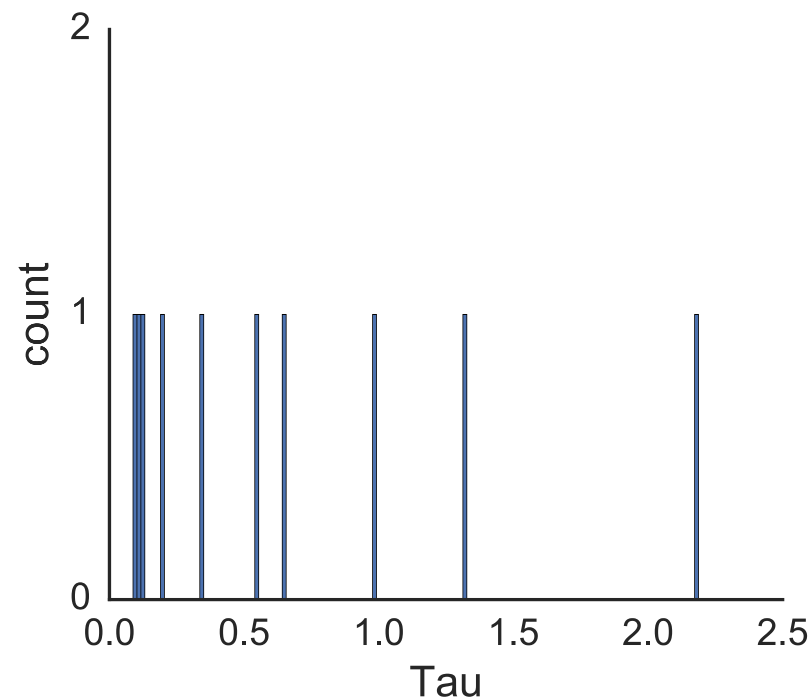
a statistical summary of the data

be equal to or more extreme than its observed value

$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total } \#)+1}$$

Calculating p-values from simulated data

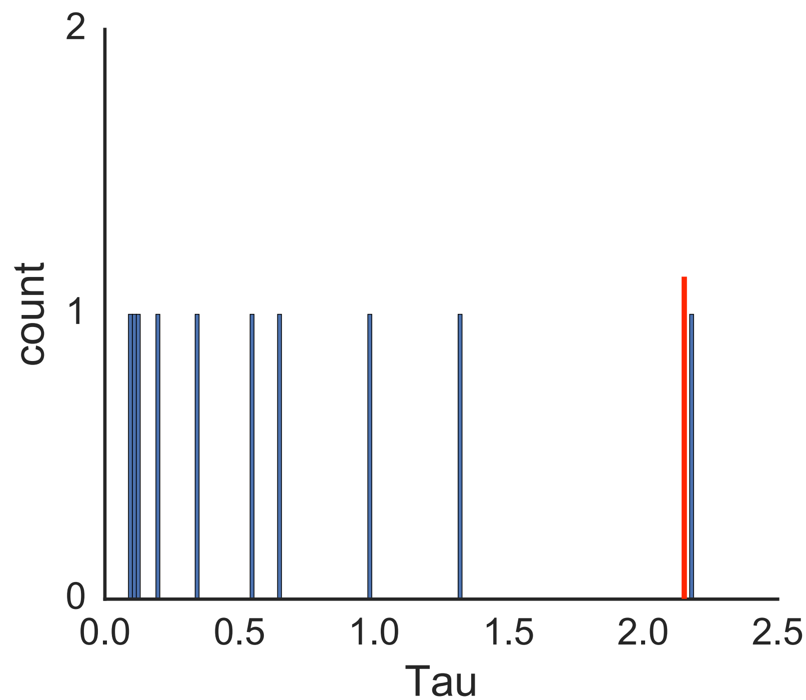
Order of p-values	p-value
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total }\#)+1}$$

Calculating p-values from simulated data

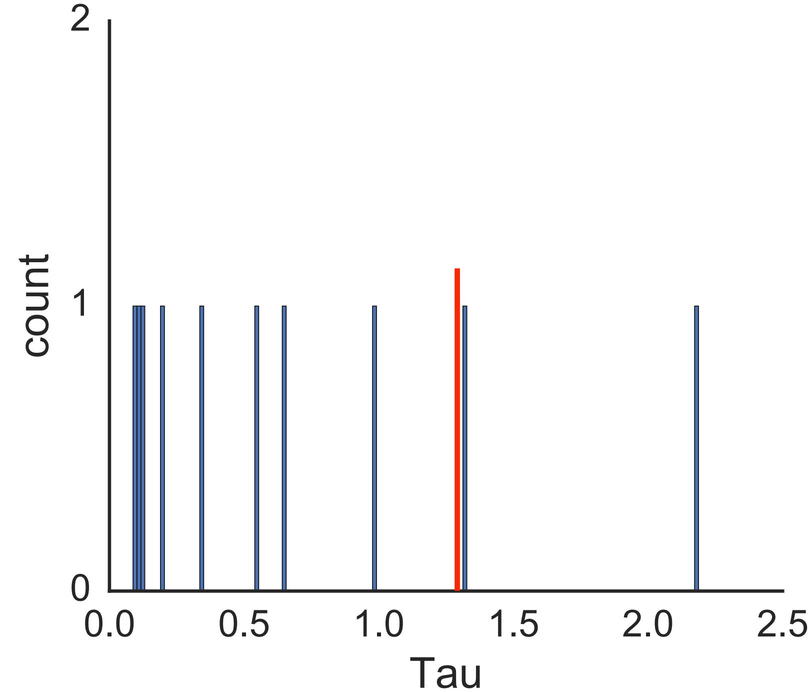
Order of p-values	p-value
1	(1+1)/(10+1) = 0.18
2	
3	
4	
5	
6	
7	
8	
9	
10	



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total }\#)+1}$$

Generating p-values

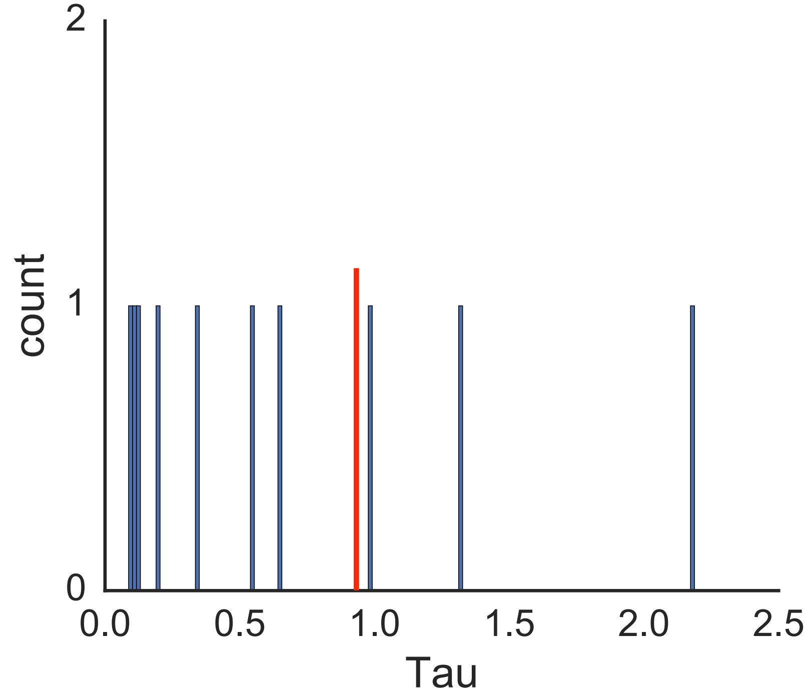
Order of p-values	p-value
1	0.18
2	(2+1)/(10+1) = 0.27
3	
4	
5	
6	
7	
8	
9	
10	



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total }\#)+1}$$

Generating p-values

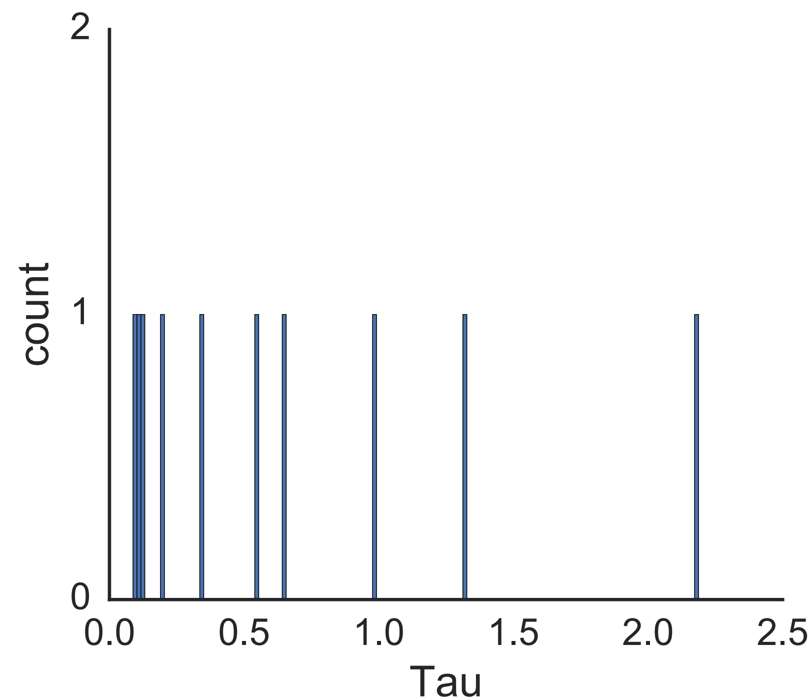
Order of p-values	p-value
1	0.18
2	0.27
3	(3+1)/(10+1) = 0.36
4	
5	
6	
7	
8	
9	
10	



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total }\#)+1}$$

Generating p-values

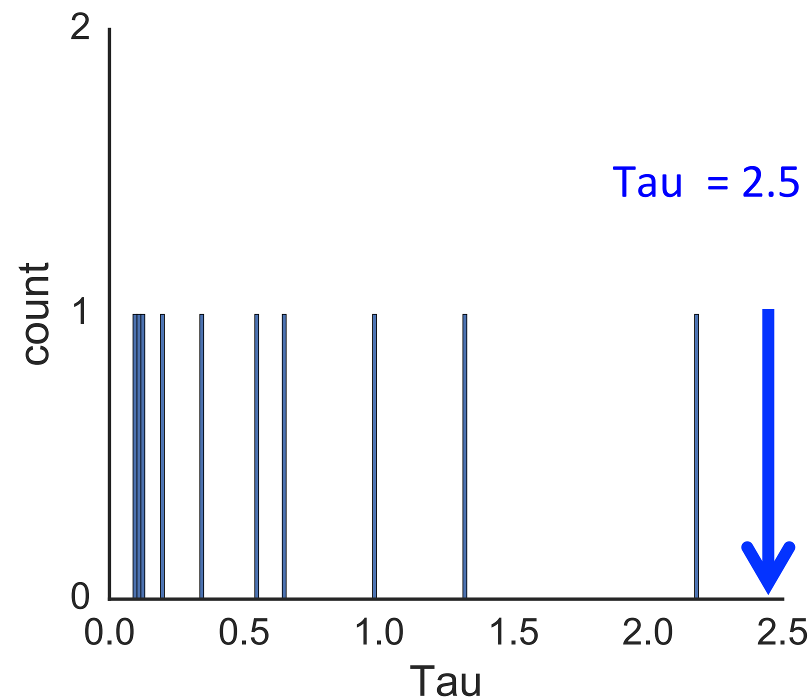
Order of p-values	p-value
1	0.18
2	0.27
3	0.36
4	0.45
5	0.55
6	0.64
7	0.73
8	0.82
9	0.91
10	1.00



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total }\#)+1}$$

Compare experimental results to null distribution to generate p-value

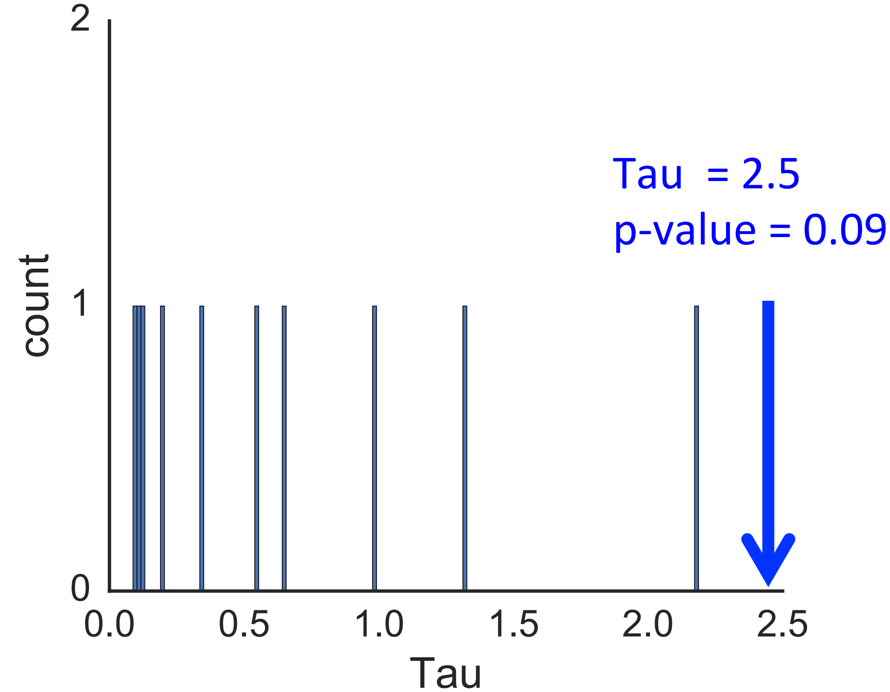
Order of p-values	p-value
1	0.18
2	0.27
3	0.36
4	0.45
5	0.55
6	0.64
7	0.73
8	0.82
9	0.91
10	1.00



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total } \#)+1}$$

Compare experimental results to null distribution to generate p-value

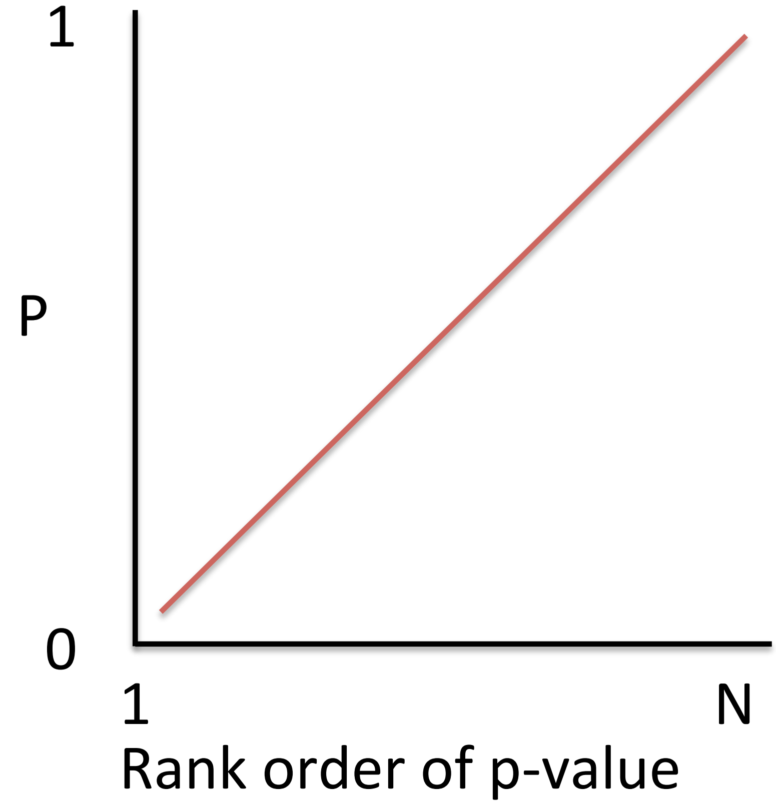
Order of p-values	p-value
1	0.18
2	0.27
3	0.36
4	0.45
5	0.55
6	0.64
7	0.73
8	0.82
9	0.91
10	1.00



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total } \#)+1}$$

P-values are uniform under the null distribution

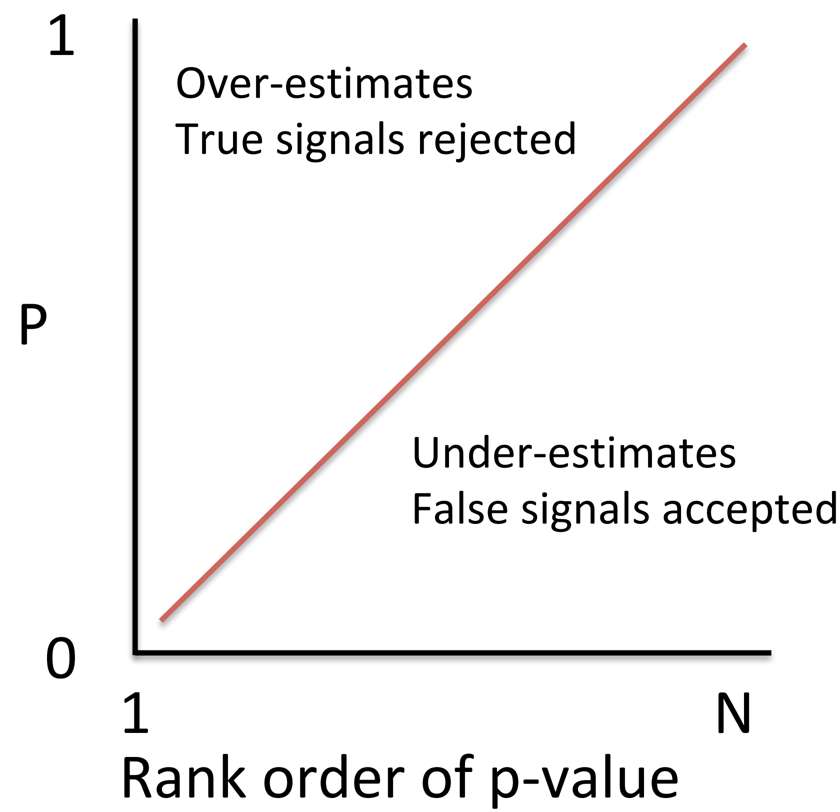
Order of p-values	p-value
1	0.18
2	0.27
3	0.36
4	0.45
5	0.55
6	0.64
7	0.73
8	0.82
9	0.91
10	1.00



$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total }\#)+1}$$

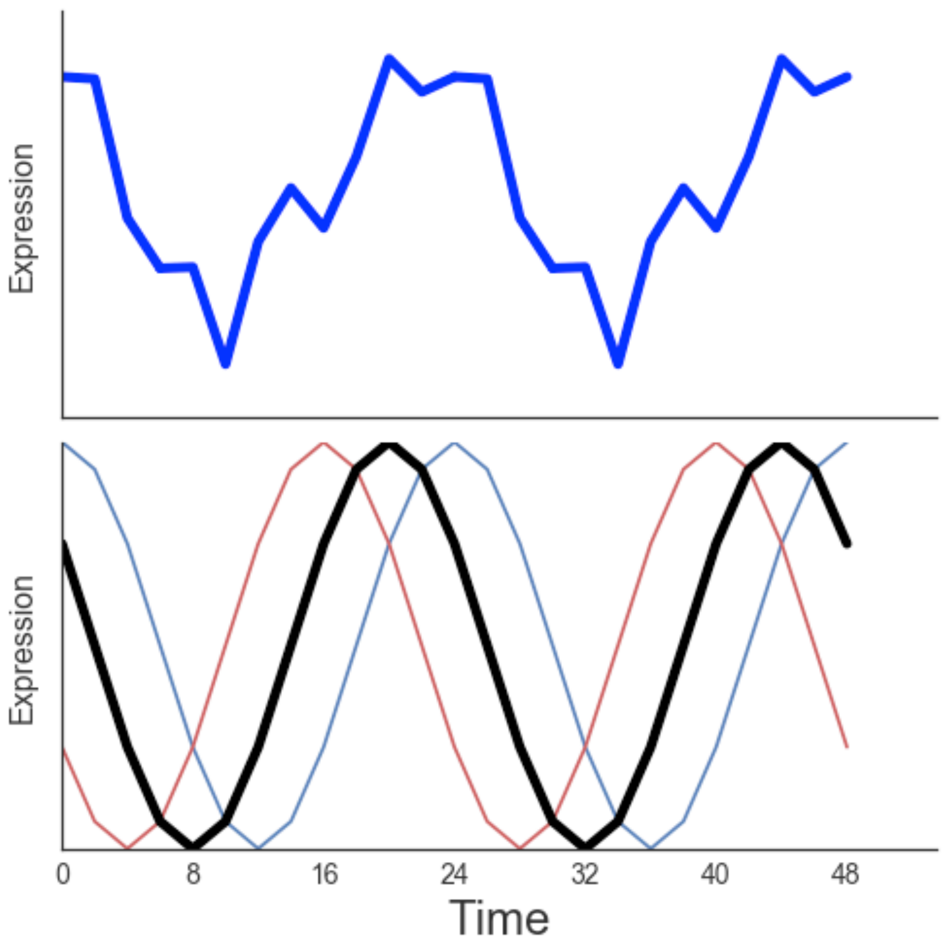
P-values are uniform under the null distribution

Order of p-values	p-value
1	0.18
2	0.27
3	0.36
4	0.45
5	0.55
6	0.64
7	0.73
8	0.82
9	0.91
10	1.00



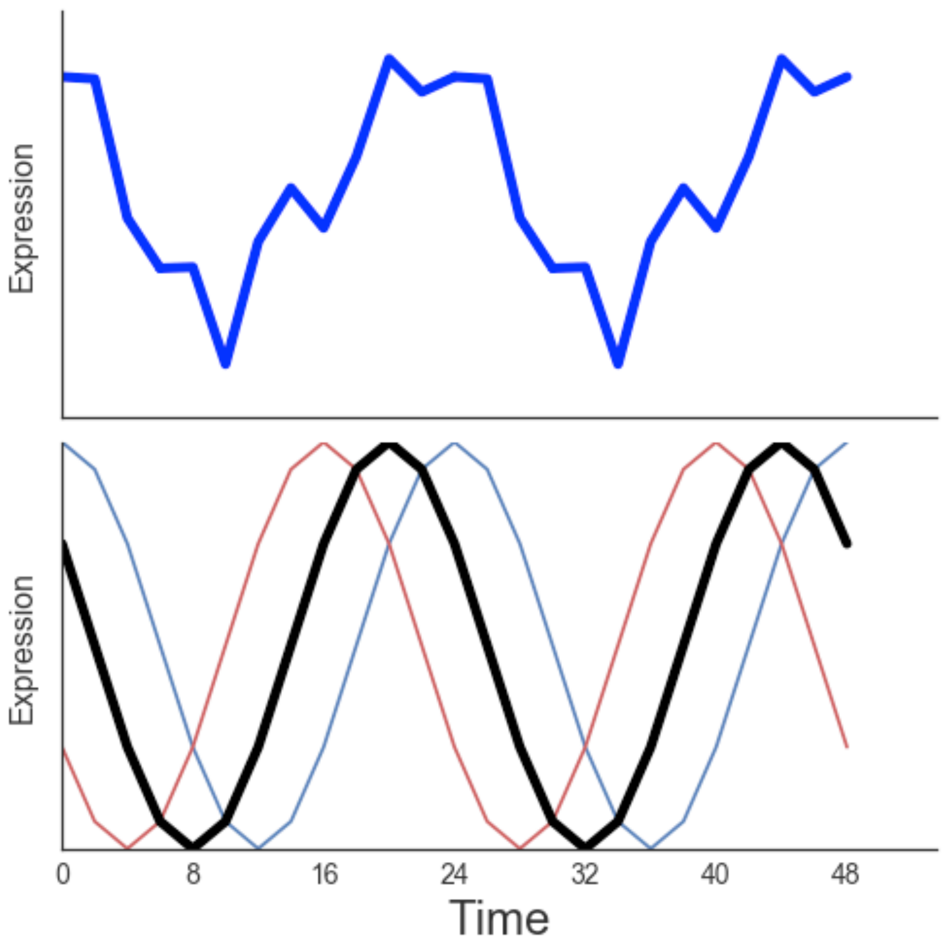
$$\text{p-value} = \frac{(\# \geq \text{observed value})+1}{(\text{Total } \#) + 1}$$

Kendall Tau p-values underestimate the true p-values

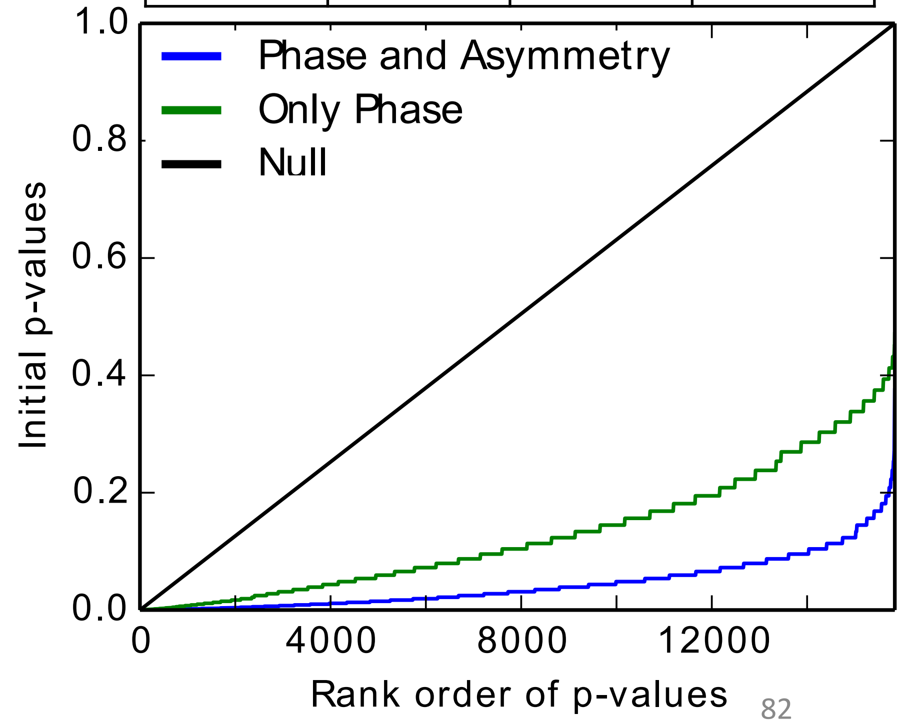


Period	Phase	τ	p-value
24	4	0.4	0.3
24	**8**	**1.1**	**0.001**
24	12	0.8	0.02

Kendall Tau p-values underestimate the true p-values

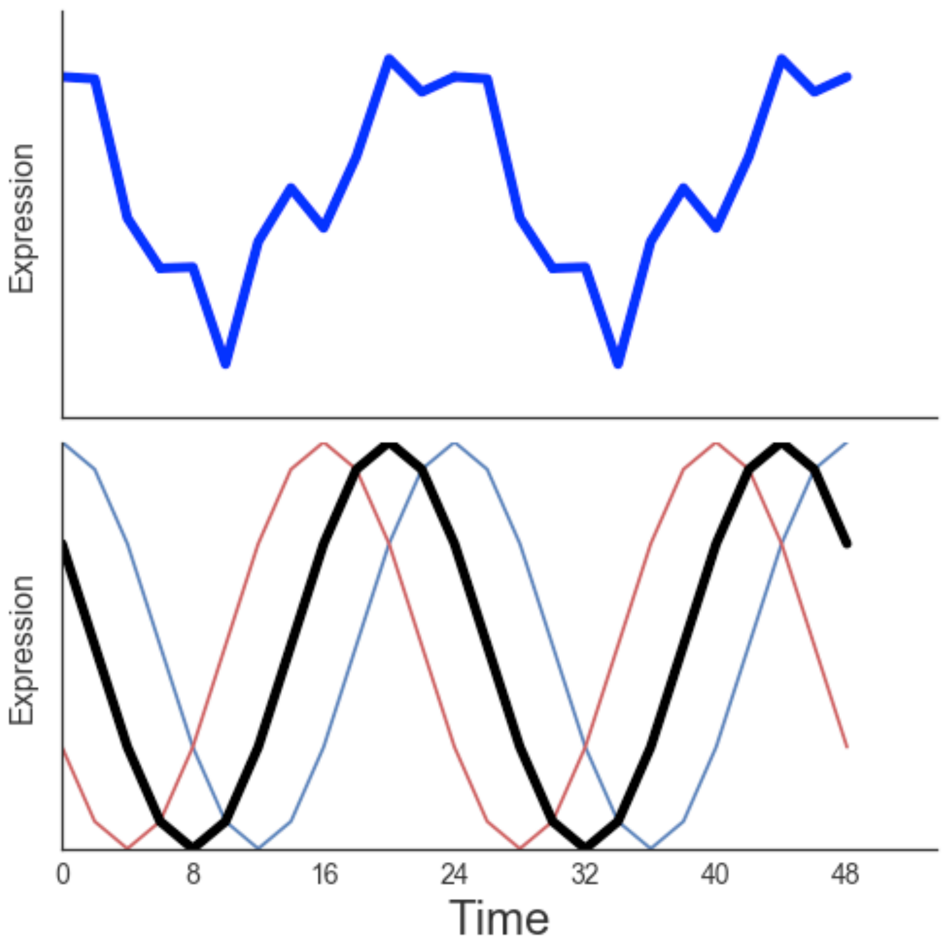


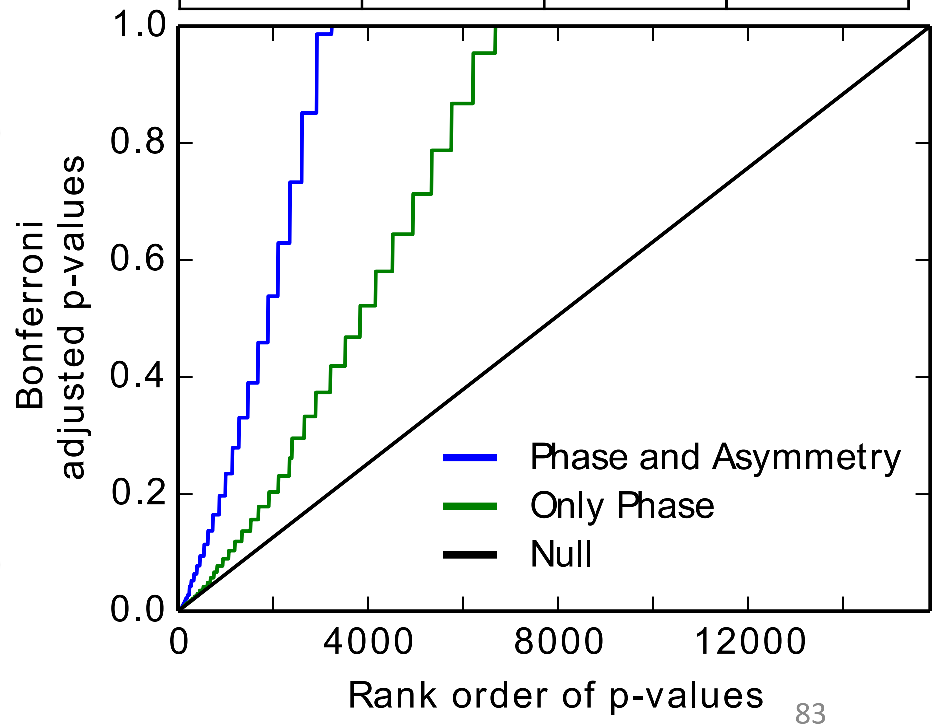
Period	Phase	τ	p-value
24	4	0.4	0.3
24	**8**	**1.1**	**0.001**
24	12	0.8	0.02



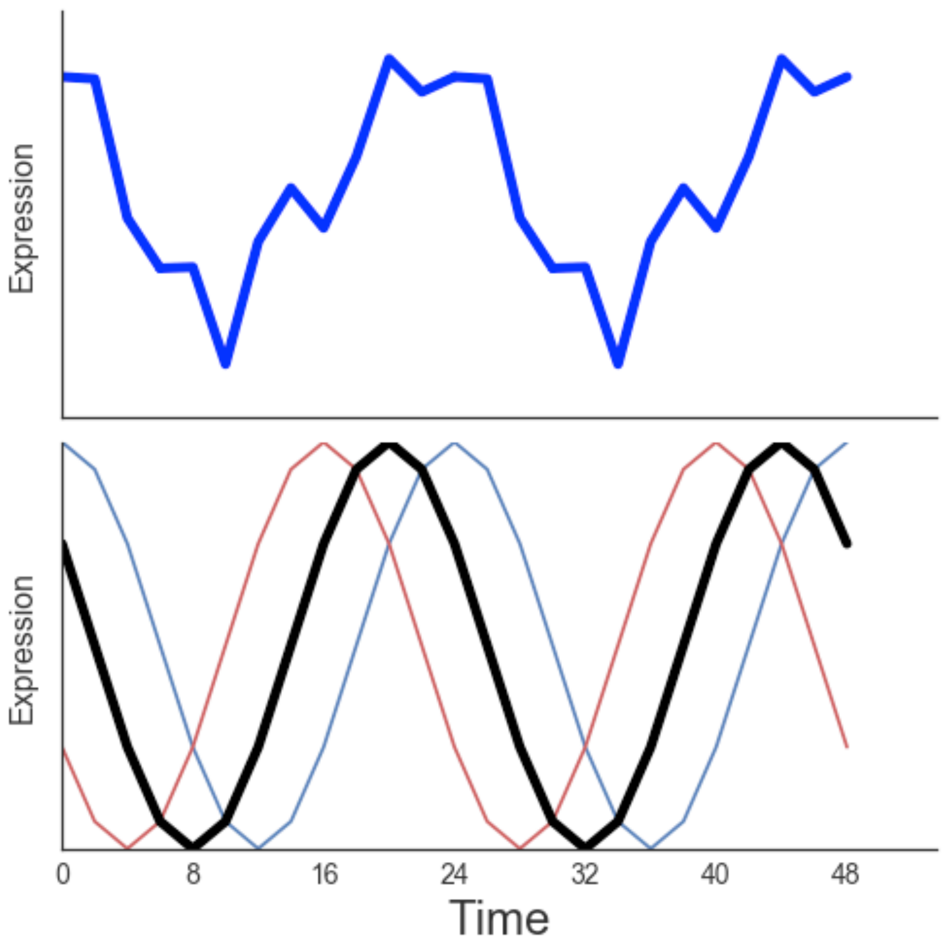
The Bonferroni correction results in overestimates of p-values

τ	p-value	# ref.	Bonf.
0.4	0.3	12	1
1.1	**0.001**	**12**	**0.012**
0.8	0.02	12	0.24

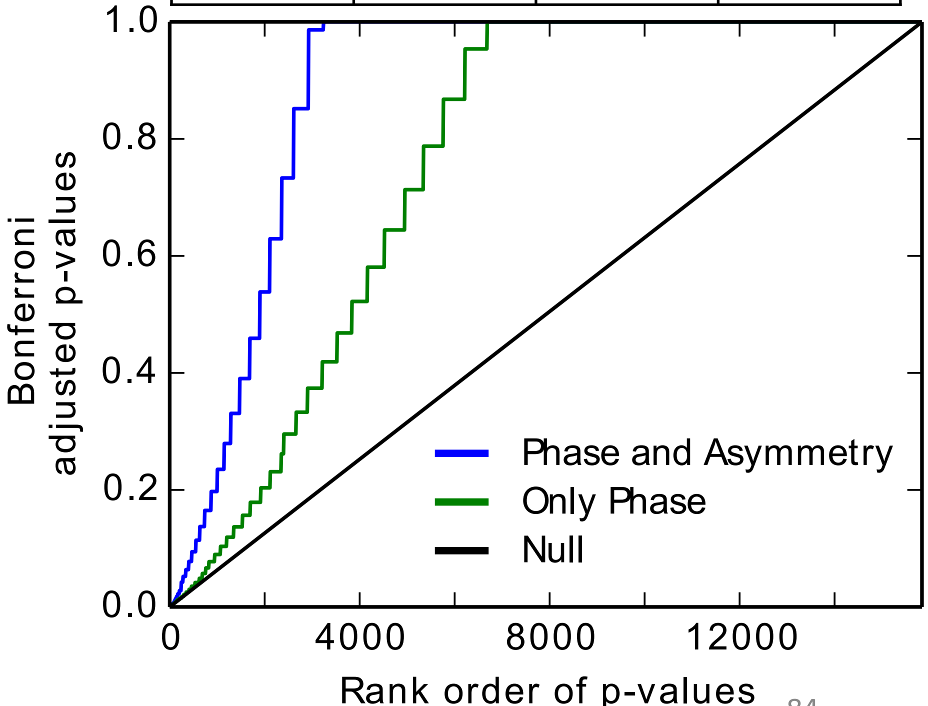




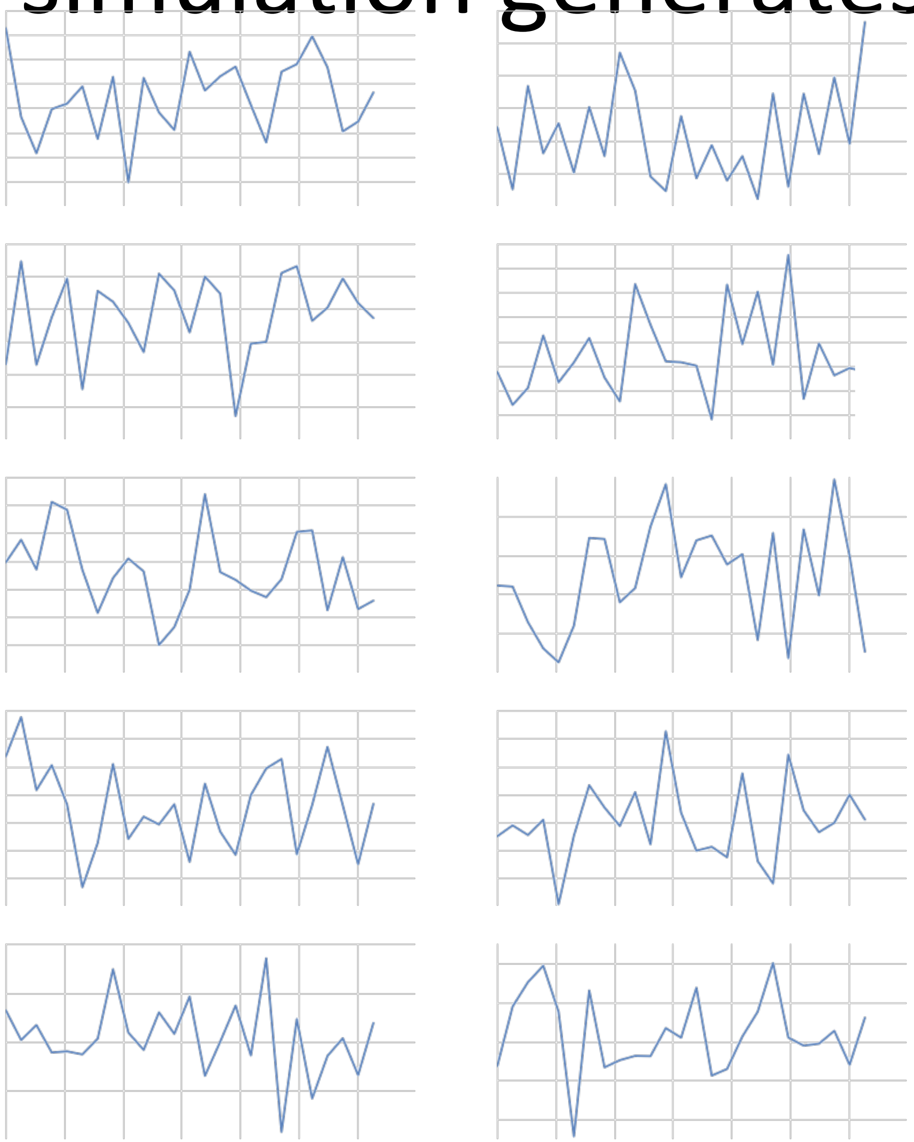
The Bonferroni correction results in overestimates of p-values



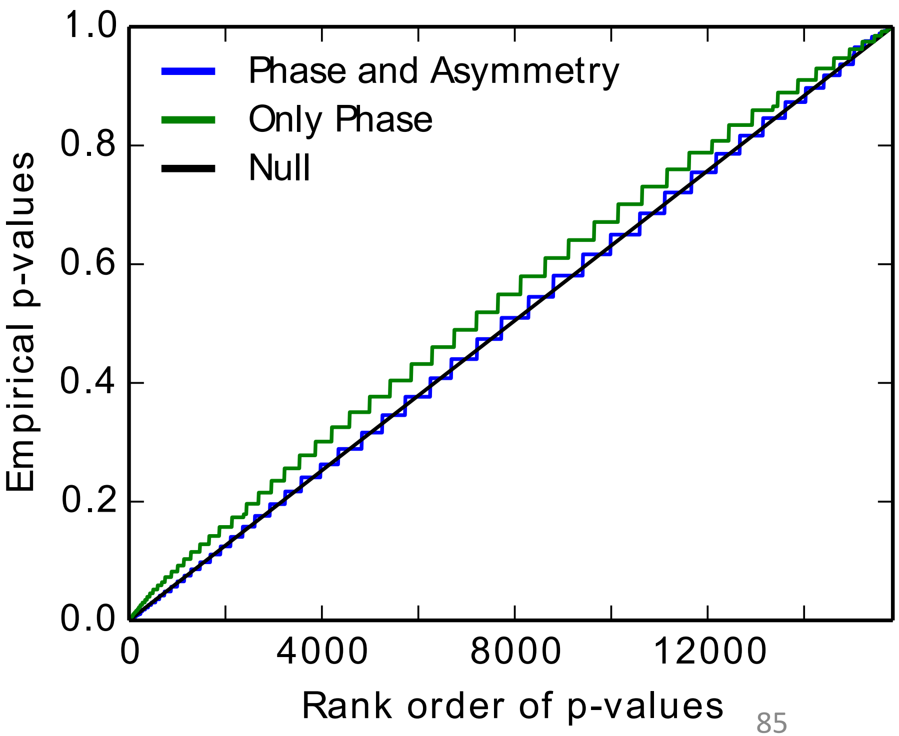
τ	p-value	# ref.	Bonf.
0.4	0.3	132	1
1.1	**0.001**	**132**	**0.132**
0.8	0.02	132	1



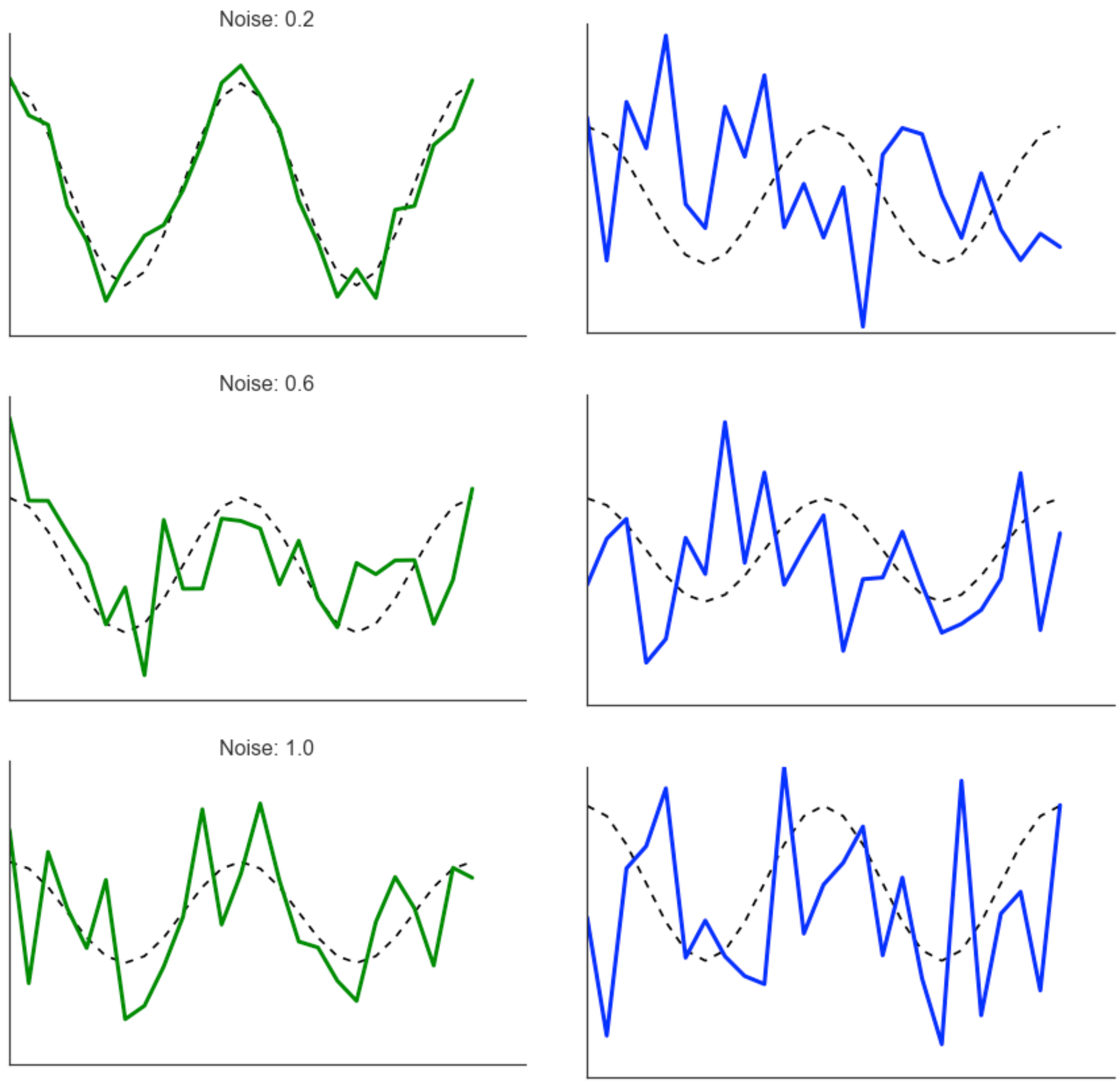
Empirically calculating the p-values via simulation generates accurate p-values



Simulate 1 million time series from noise to get empirical distribution of null p-values

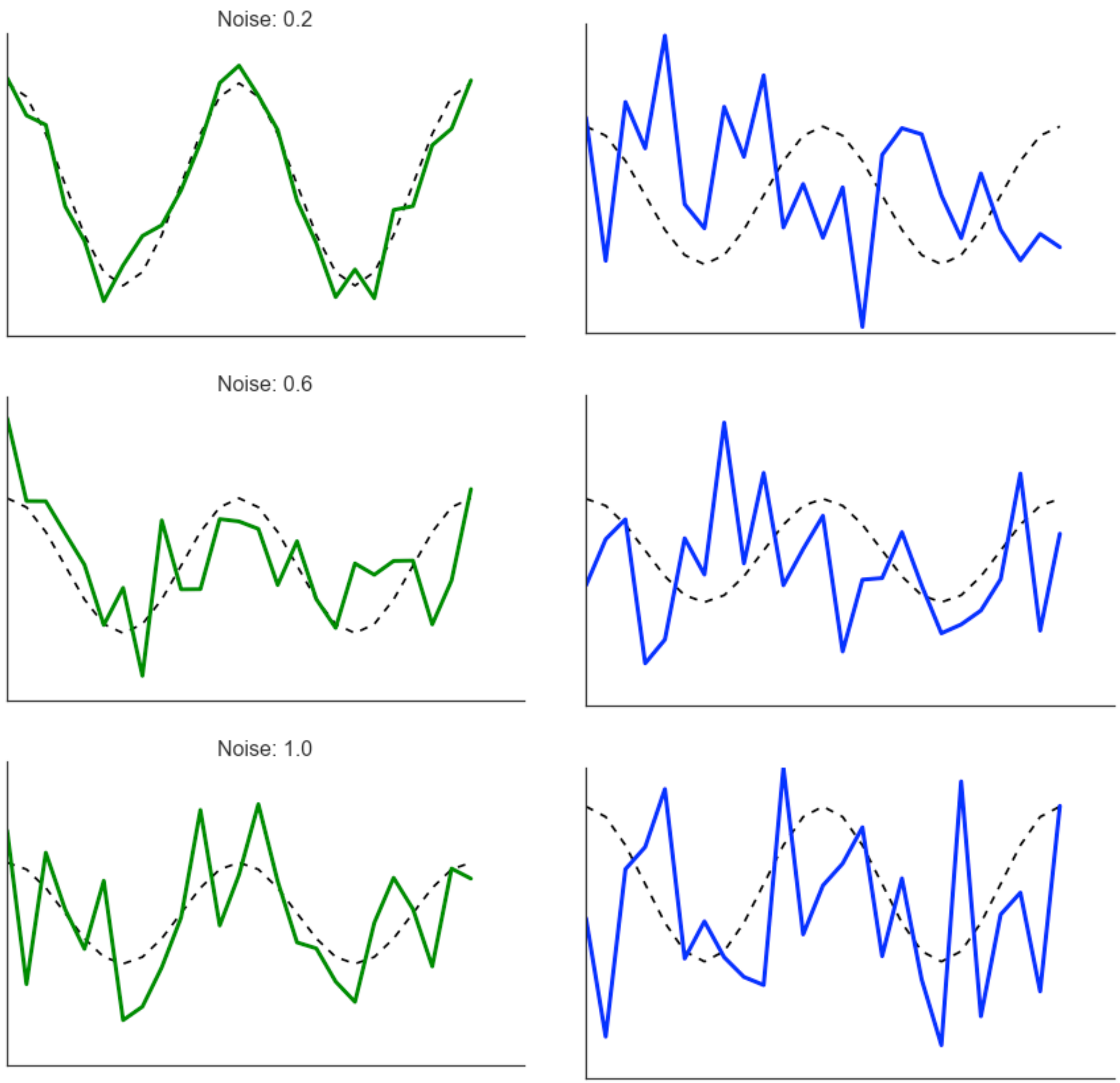


Simulated data comparison



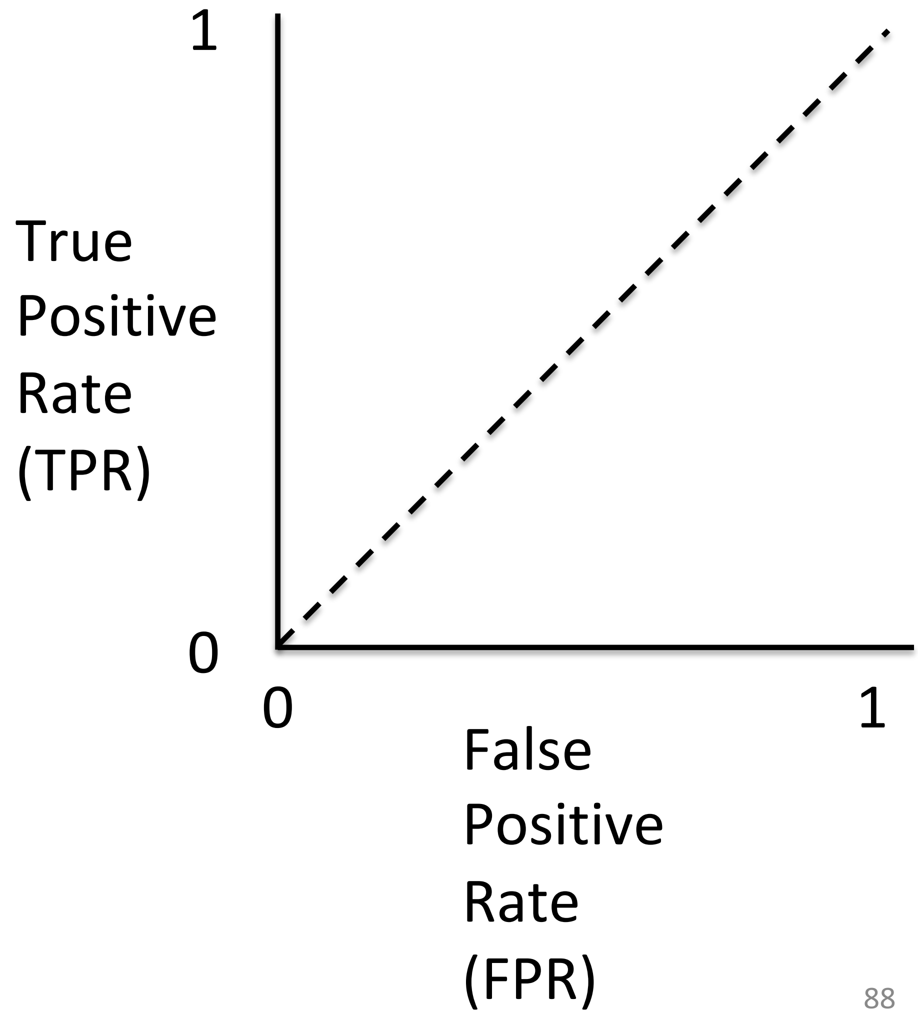
Simulated data comparison

Rhythmic?	**Rhythmic?**
Positive	Negative
Positive	Negative
Positive	Negative



Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1		Positive
B	0.2		Negative
C	0.22		Positive
D	0.3		Negative
E	0.31		Positive
F	0.5		Negative
G	0.6		Negative
H	0.78		Negative

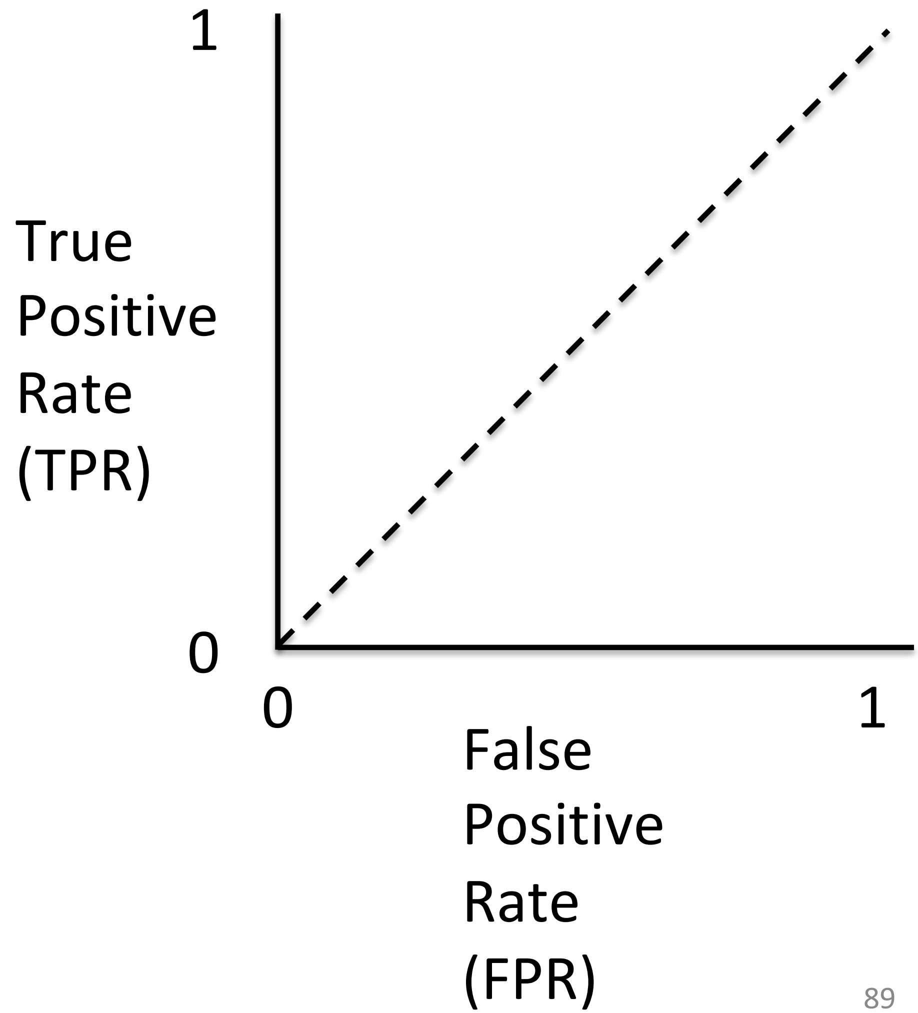


TPR = True Pos. / Positives = TP / 3

FPR = False Pos. / Negatives = FP / 5

Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1		**Positive**
B	0.2		Negative
C	0.22		**Positive**
D	0.3		Negative
E	0.31		**Positive**
F	0.5		Negative
G	0.6		Negative
H	0.78		Negative

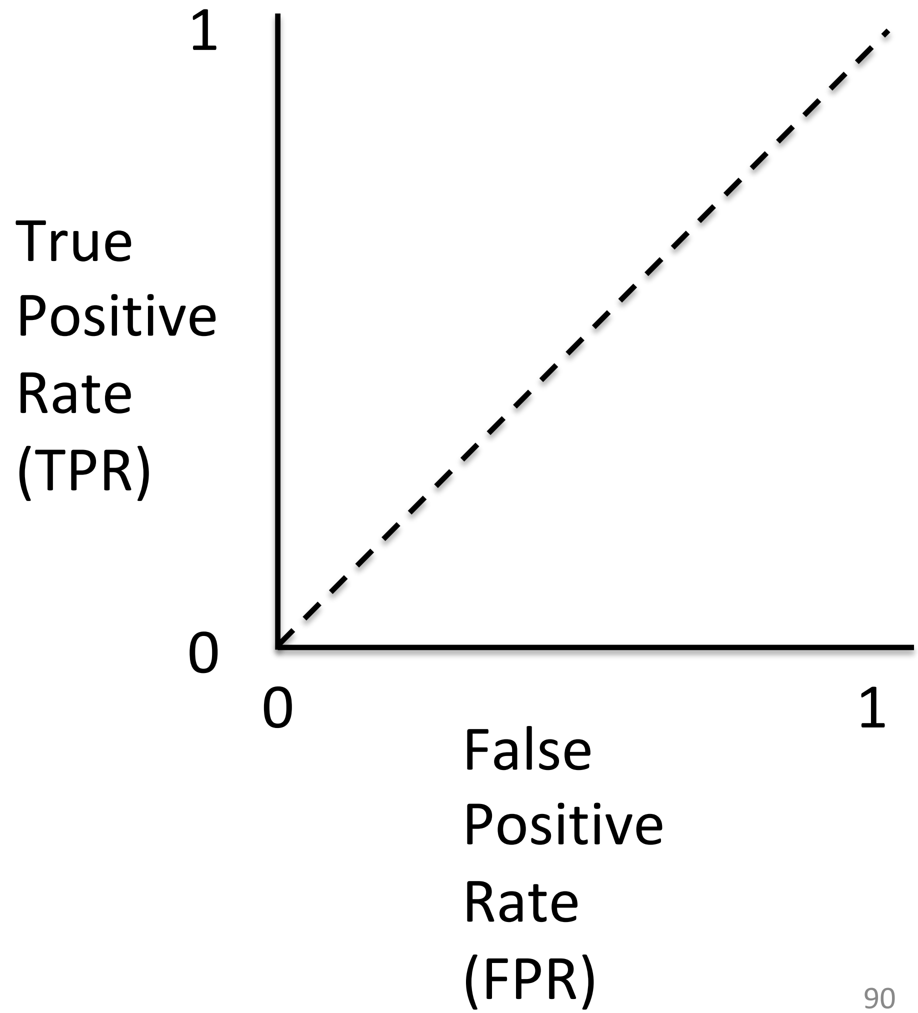


TPR = True Pos. / **Positives** = TP / **3**

FPR = False Pos. / Negatives = FP / 5

Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1		Positive
B	0.2		**Negative**
C	0.22		Positive
D	0.3		**Negative**
E	0.31		Positive
F	0.5		**Negative**
G	0.6		**Negative**
H	0.78		**Negative**



TPR = True Pos. / Positives = TP / 3

FPR = False Pos. / **Negatives** = FP / **5**

Simulated data comparison

Pos.

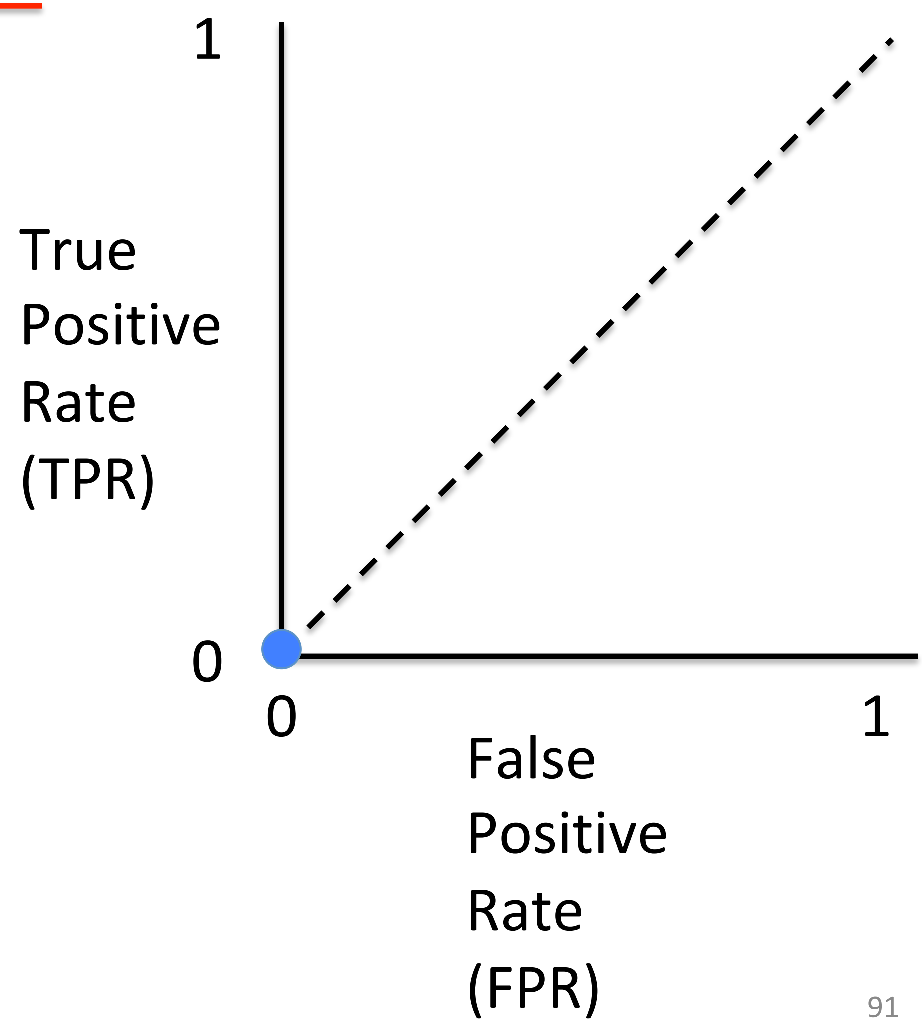
ID	Score	Classification	Rhythmic?
A	0.1	Negative	Positive
B	0.2	Negative	Negative
C	0.22	Negative	Positive
D	0.3	Negative	Negative
E	0.31	Negative	Positive
F	0.5	Negative	Negative
G	0.6	Negative	Negative
H	0.78	Negative	Negative

Neg.

Threshold: 0.05

TPR = True Pos. / Positives = 0 / 3

FPR = False Pos. / Negatives = 0 / 5



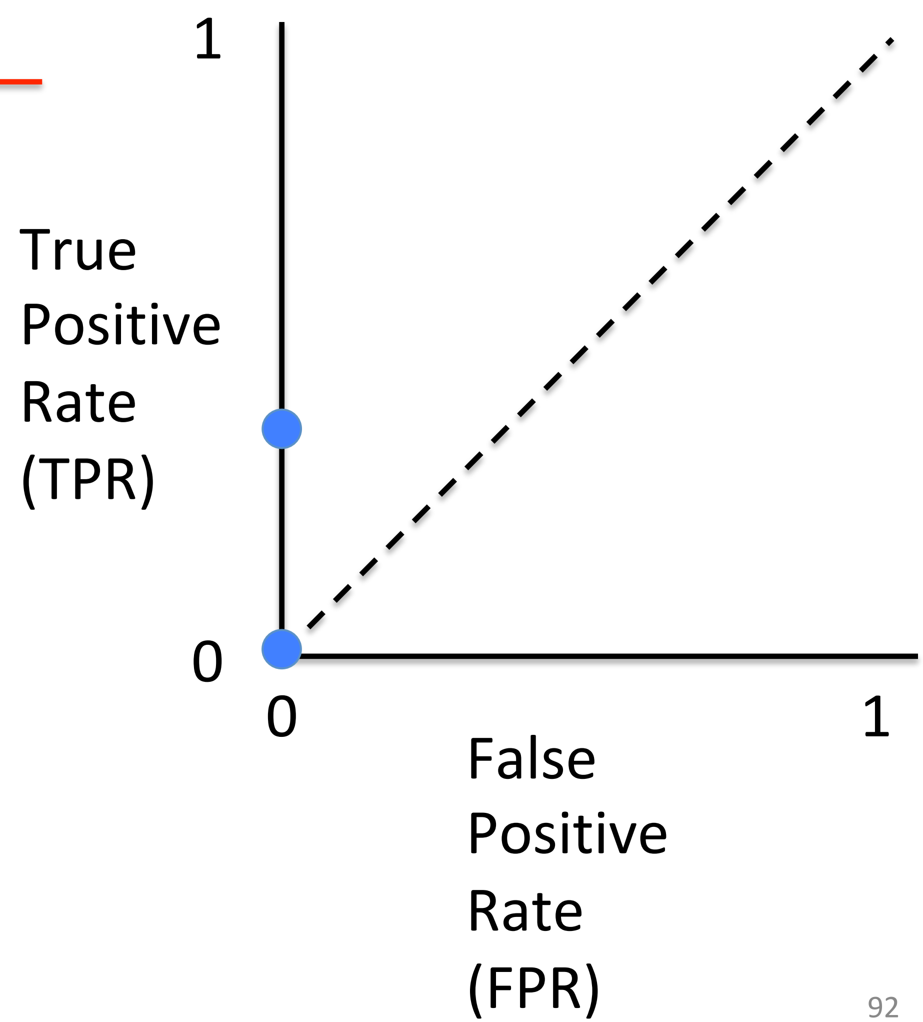
Simulated data comparison

	ID	Score	Classification	Rhythmic?
Pos.	A	0.1	Positive	Positive
Neg.	B	0.2	Negative	Negative
	C	0.22	Negative	Positive
	D	0.3	Negative	Negative
	E	0.31	Negative	Positive
	F	0.5	Negative	Negative
	G	0.6	Negative	Negative
	H	0.78	Negative	Negative

Threshold: 0.1

TPR = True Pos. / Positives = 1 / 3

FPR = False Pos. / Negatives = 0 / 5



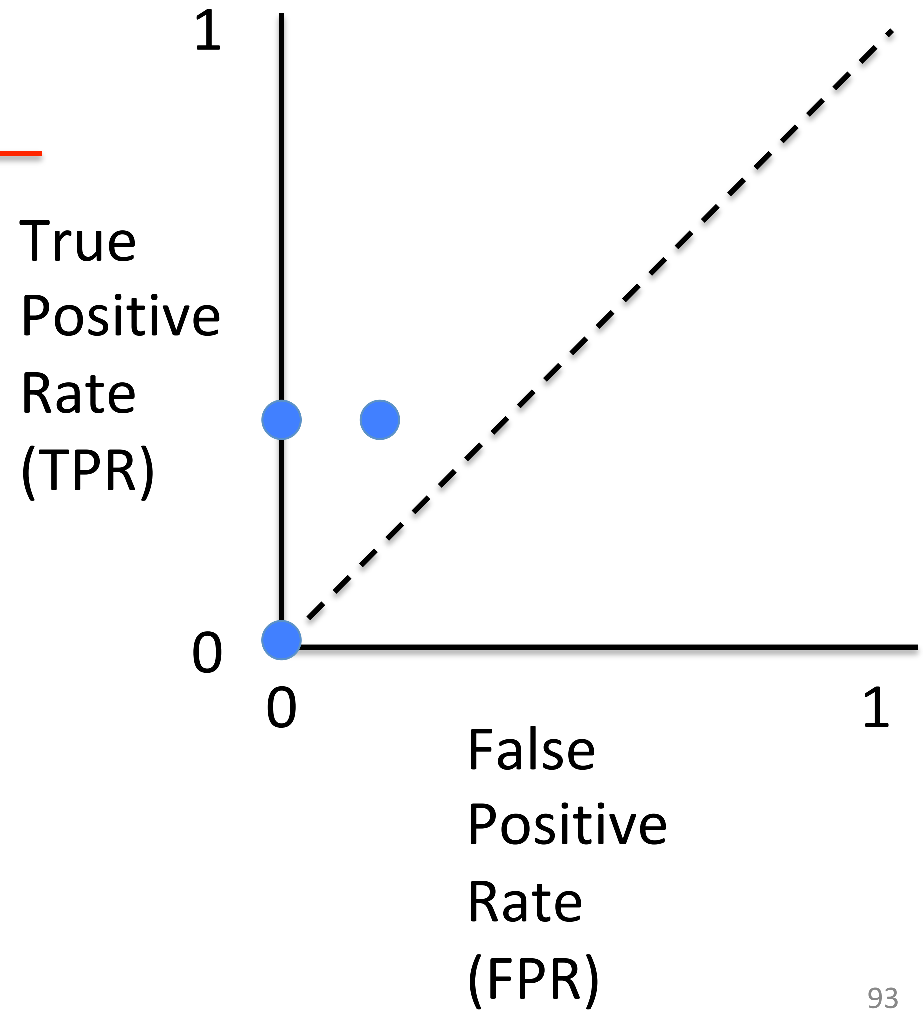
Simulated data comparison

	ID	Score	Classification	Rhythmic?
Pos.	A	0.1	Positive	Positive
	B	0.2	Positive	Negative
Neg.	C	0.22	Negative	Positive
	D	0.3	Negative	Negative
	E	0.31	Negative	Positive
	F	0.5	Negative	Negative
	G	0.6	Negative	Negative
	H	0.78	Negative	Negative

Threshold: 0.2

TPR = True Pos. / Positives = 1 / 3

FPR = False Pos. / Negatives = 1 / 5



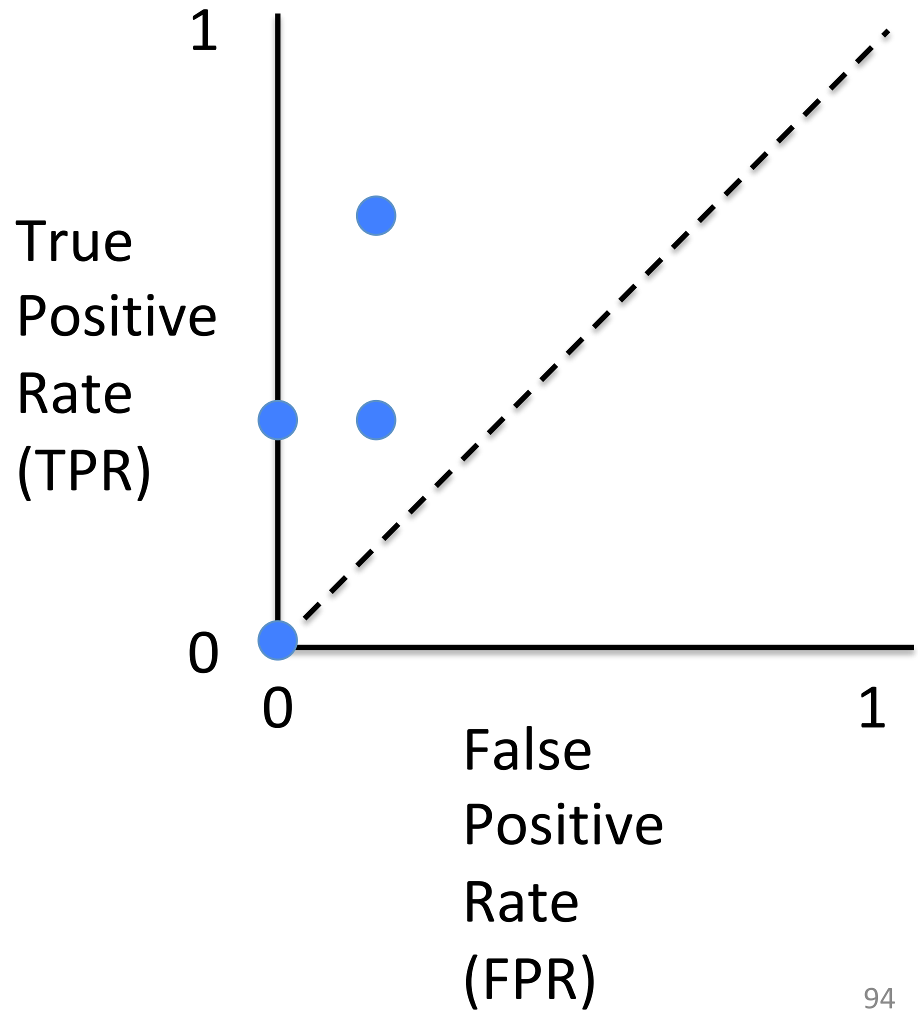
Simulated data comparison

	ID	Score	Classification	Rhythmic?
	A	0.1	Positive	Positive
Pos.	B	0.2	Positive	Negative
	C	0.22	Positive	Positive
	D	0.3	Negative	Negative
Neg.	E	0.31	Negative	Positive
	F	0.5	Negative	Negative
	G	0.6	Negative	Negative
	H	0.78	Negative	Negative

Threshold: 0.22

TPR = True Pos. / Positives = 2 / 3

FPR = False Pos. / Negatives = 1 / 5



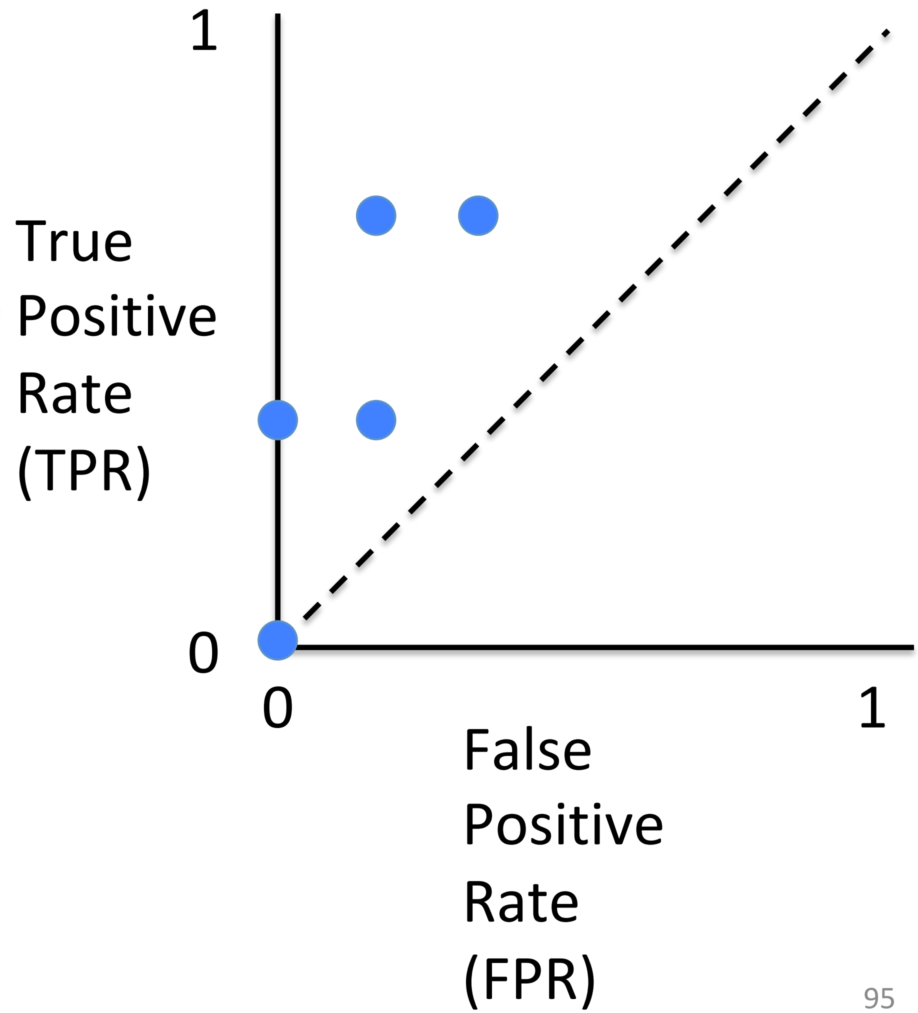
Simulated data comparison

	ID	Score	Classification	Rhythmic?
	A	0.1	Positive	Positive
	B	0.2	Positive	Negative
Pos.	C	0.22	Positive	Positive
	D	0.3	Positive	Negative
Neg.	E	0.31	Negative	Positive
	F	0.5	Negative	Negative
	G	0.6	Negative	Negative
	H	0.78	Negative	Negative

Threshold: 0.3

TPR = True Pos. / Positives = 2 / 3

FPR = False Pos. / Negatives = 2 / 5



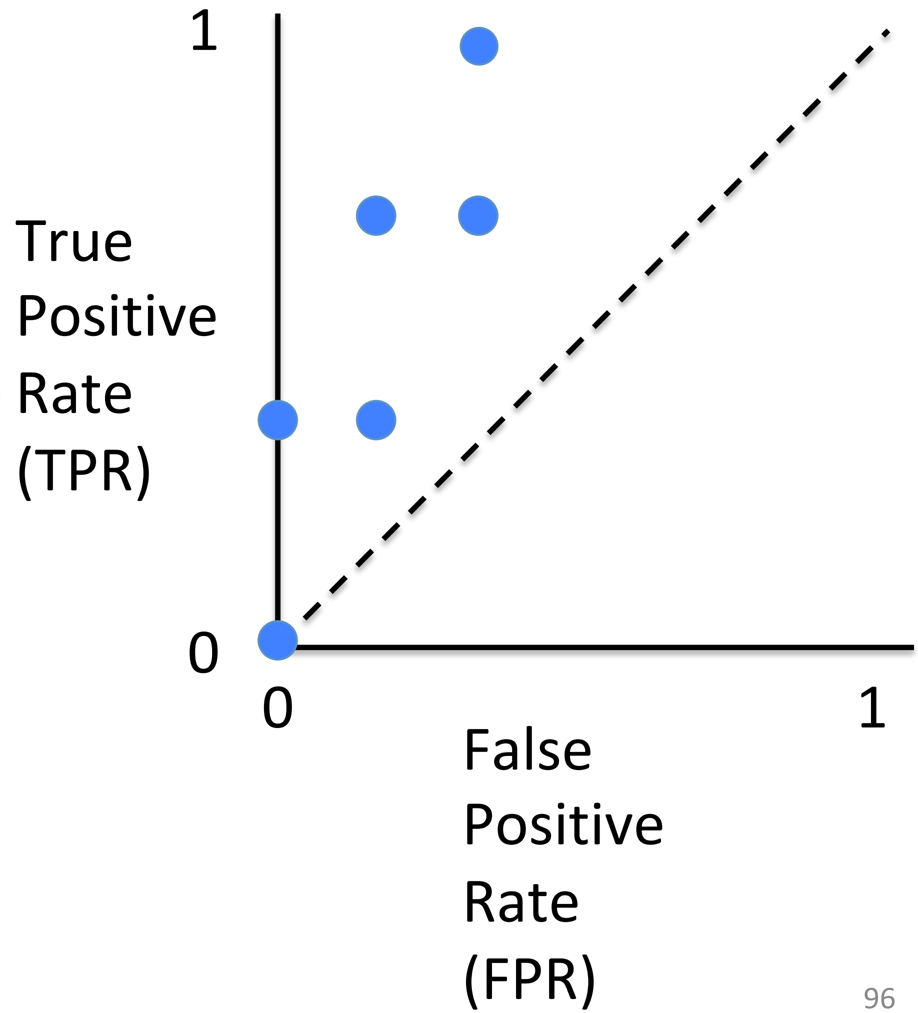
Simulated data comparison

	ID	Score	Classification	Rhythmic?
	A	0.1	Positive	Positive
	B	0.2	Positive	Negative
	C	0.22	Positive	Positive
Pos.	D	0.3	Positive	Negative
	E	0.31	Positive	Positive
Neg.	F	0.5	Negative	Negative
	G	0.6	Negative	Negative
	H	0.78	Negative	Negative

Threshold: 0.31

TPR = True Pos. / Positives = 3 / 3

FPR = False Pos. / Negatives = 2 / 5



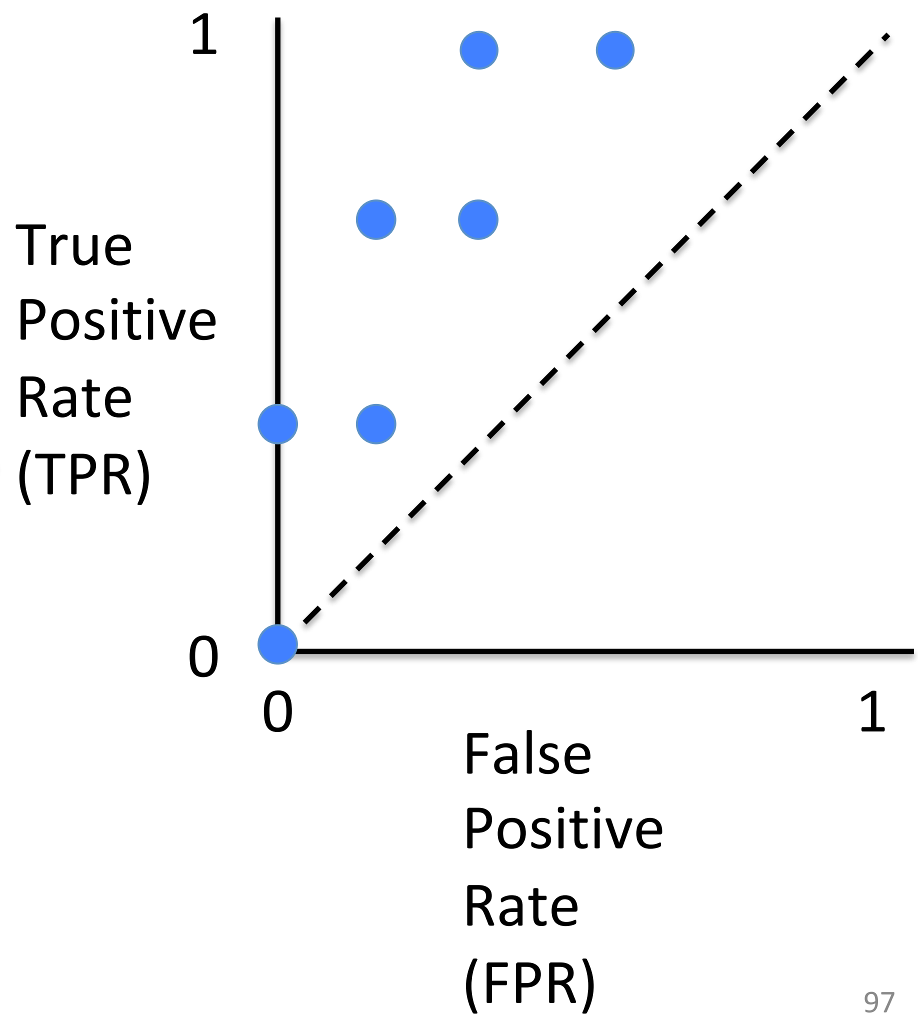
Simulated data comparison

	ID	Score	Classification	Rhythmic?
	A	0.1	Positive	Positive
	B	0.2	Positive	Negative
	C	0.22	Positive	Positive
	D	0.3	Positive	Negative
Pos.	E	0.31	Positive	Positive
	F	0.5	Positive	Negative
Neg.	G	0.6	Negative	Negative
	H	0.78	Negative	Negative

Threshold: 0.5

TPR = True Pos. / Positives = 3 / 3

FPR = False Pos. / Negatives = 3 / 5



Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1	Positive	Positive
B	0.2	Positive	Negative
C	0.22	Positive	Positive
D	0.3	Positive	Negative
E	0.31	Positive	Positive
F	0.5	Positive	Negative
G	0.6	Positive	Negative
H	0.78	Negative	Negative

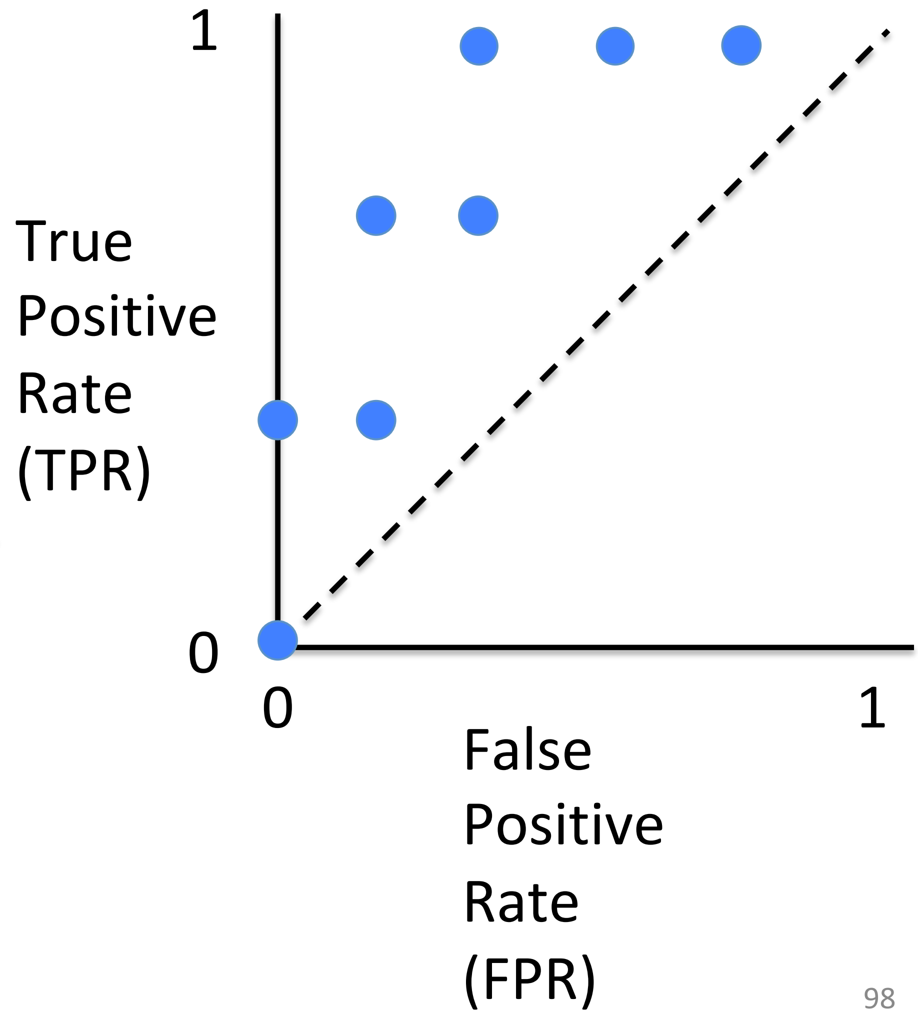
Pos.

Neg.

Threshold: 0.6

TPR = True Pos. / Positives = 3 / 3

FPR = False Pos. / Negatives = 4 / 5



Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1	Positive	Positive
B	0.2	Positive	Negative
C	0.22	Positive	Positive
D	0.3	Positive	Negative
E	0.31	Positive	Positive
F	0.5	Positive	Negative
G	0.6	Positive	Negative
H	0.78	Positive	Negative

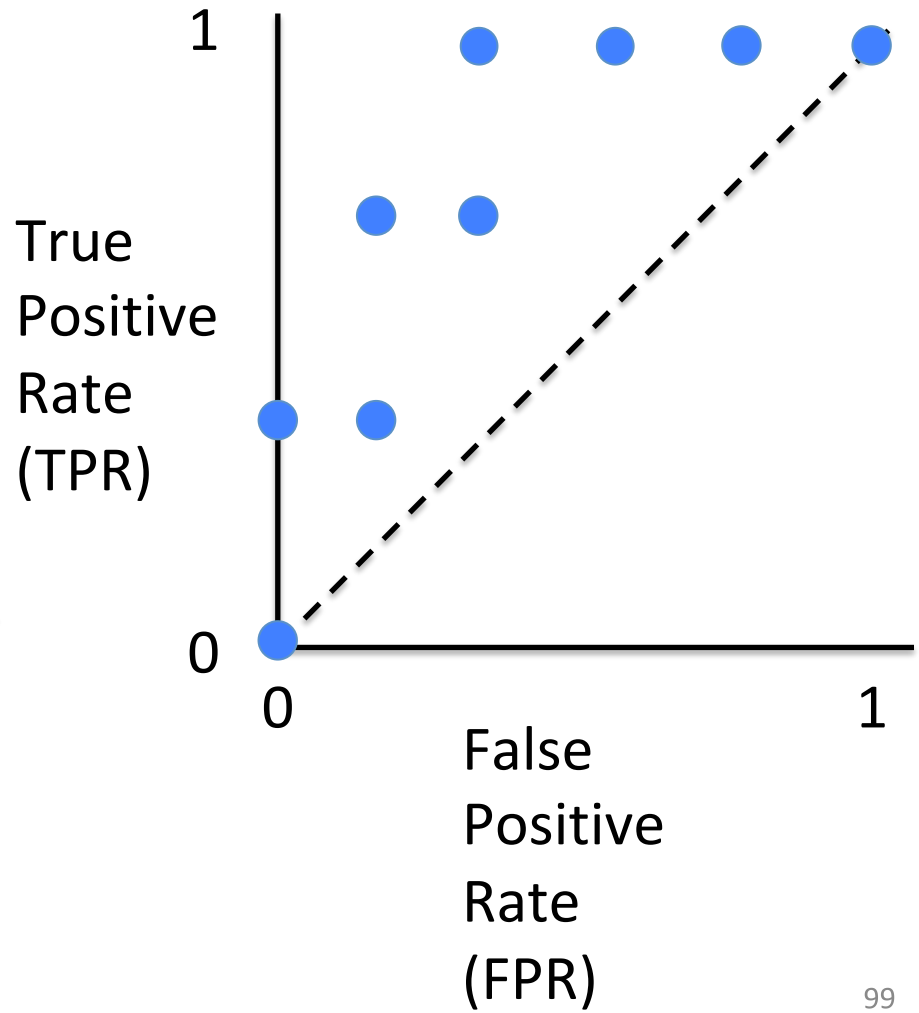
Pos.

Neg.

Threshold: 0.78

TPR = True Pos. / Positives = 3 / 3

FPR = False Pos. / Negatives = 5 / 5



Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1	Positive	Positive
B	0.2	Positive	Negative
C	0.22	Positive	Positive
D	0.3	Positive	Negative
E	0.31	Positive	Positive
F	0.5	Positive	Negative
G	0.6	Positive	Negative
H	0.78	Positive	Negative

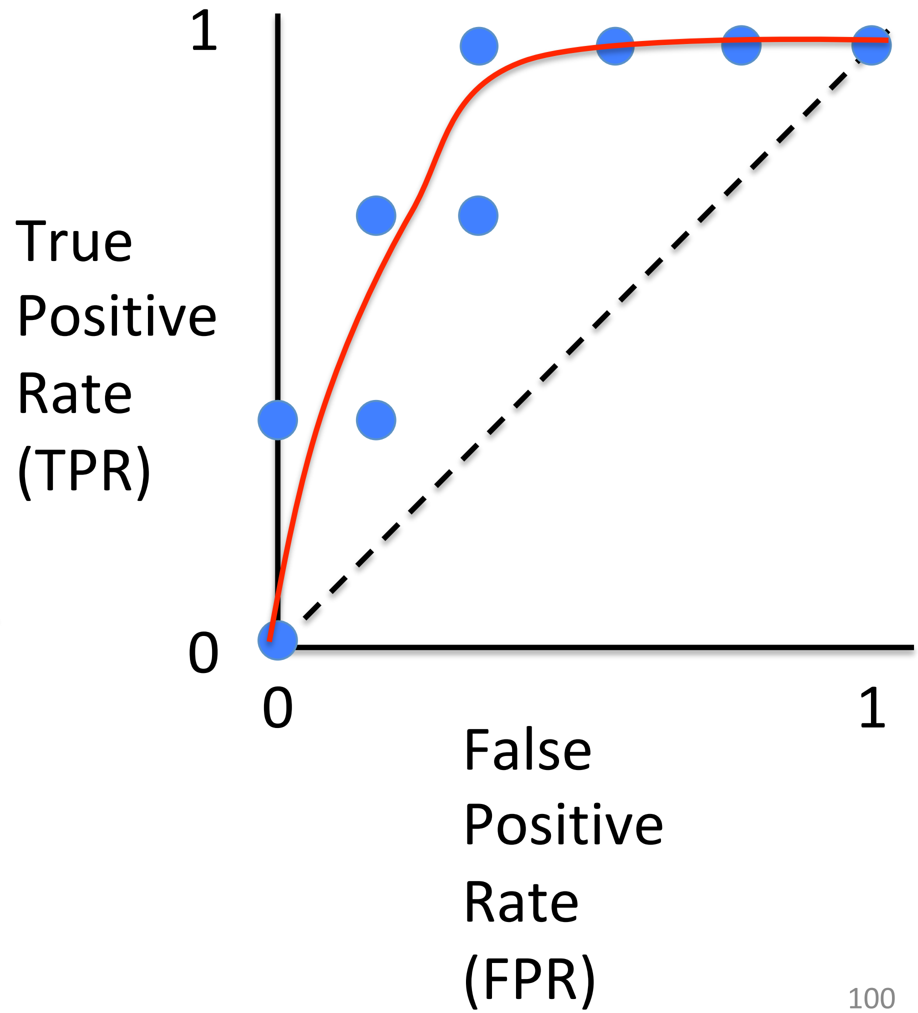
Pos.

Neg.

Threshold: 0.78

TPR = True Pos. / Positives = 3 / 3

FPR = False Pos. / Negatives = 5 / 5



Simulated data comparison

ID	Score	Classification	Rhythmic?
A	0.1	Positive	Positive
B	0.2	Positive	Negative
C	0.22	Positive	Positive
D	0.3	Positive	Negative
E	0.31	Positive	Positive
F	0.5	Positive	Negative
G	0.6	Positive	Negative
H	0.78	Positive	Negative

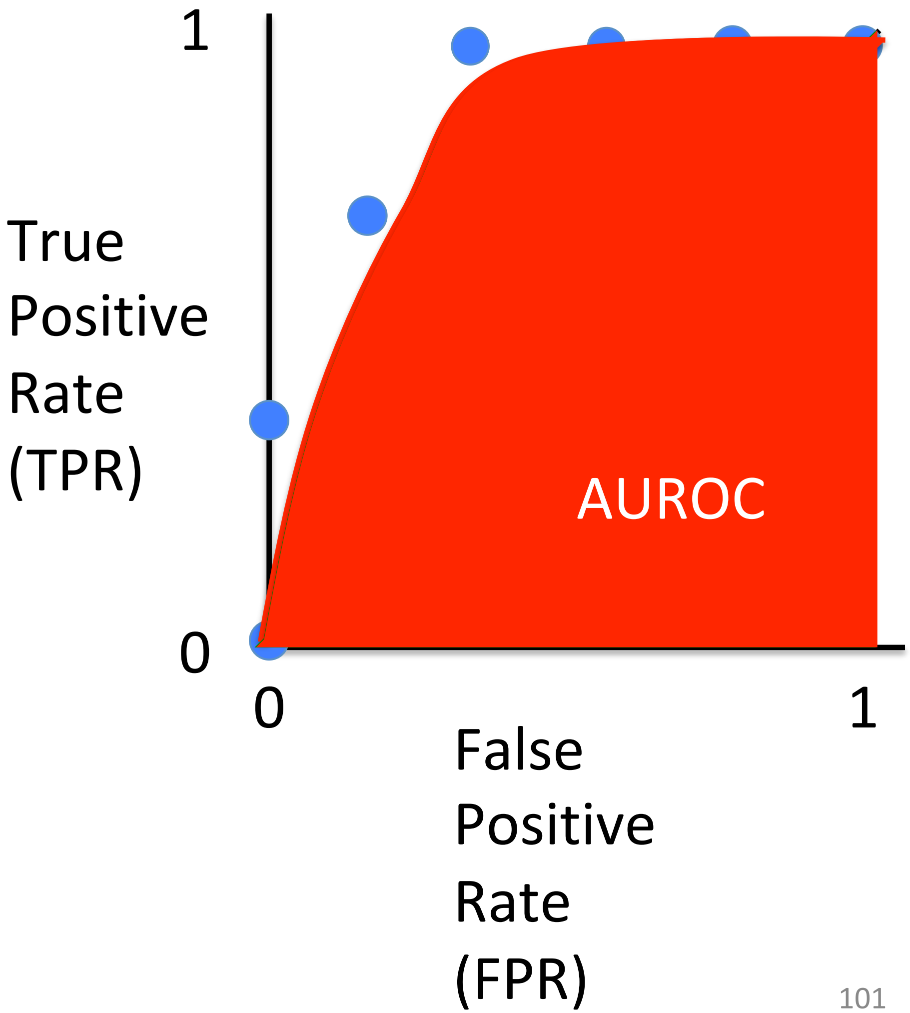
Pos.

Neg.

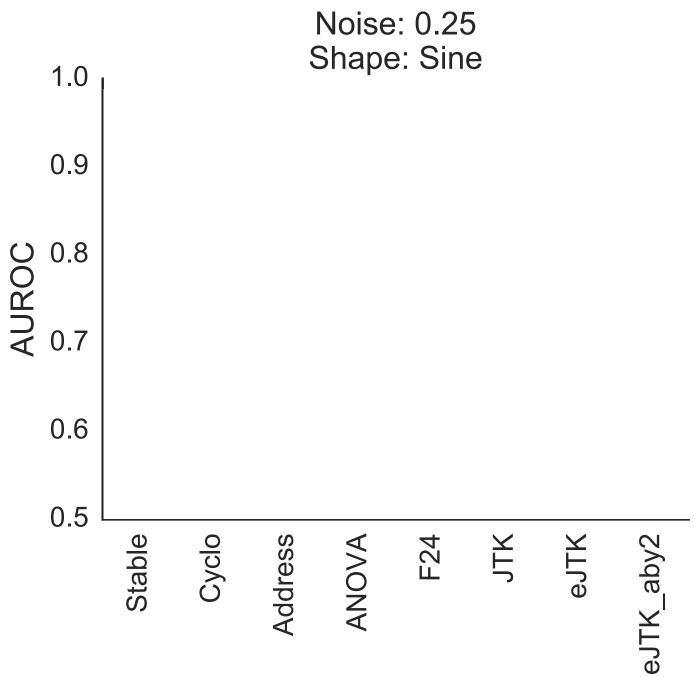
Threshold: 0.78

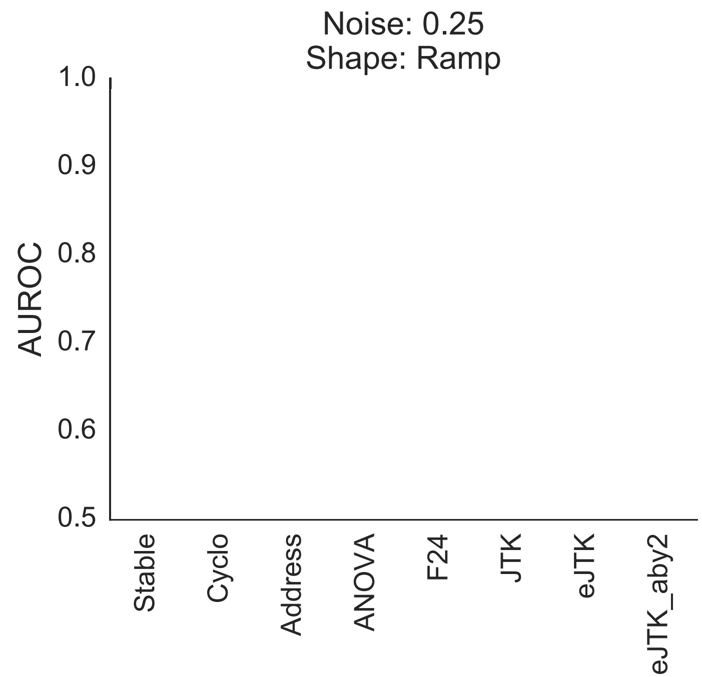
TPR = True Pos. / Positives = 3 / 3

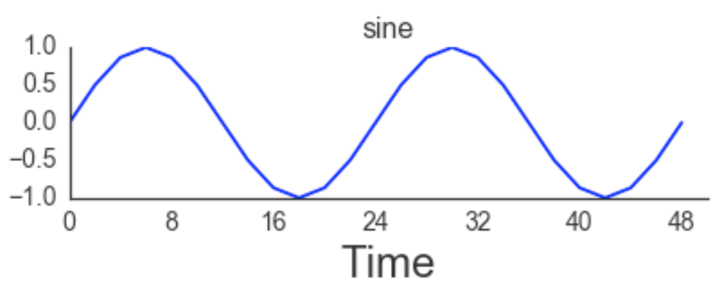
FPR = False Pos. / Negatives = 5 / 5

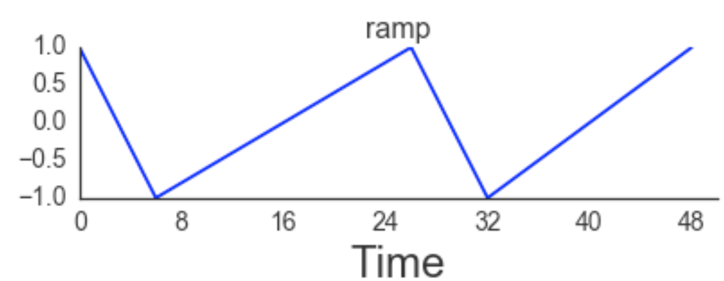


eJTK outperforms other methods on simulated data

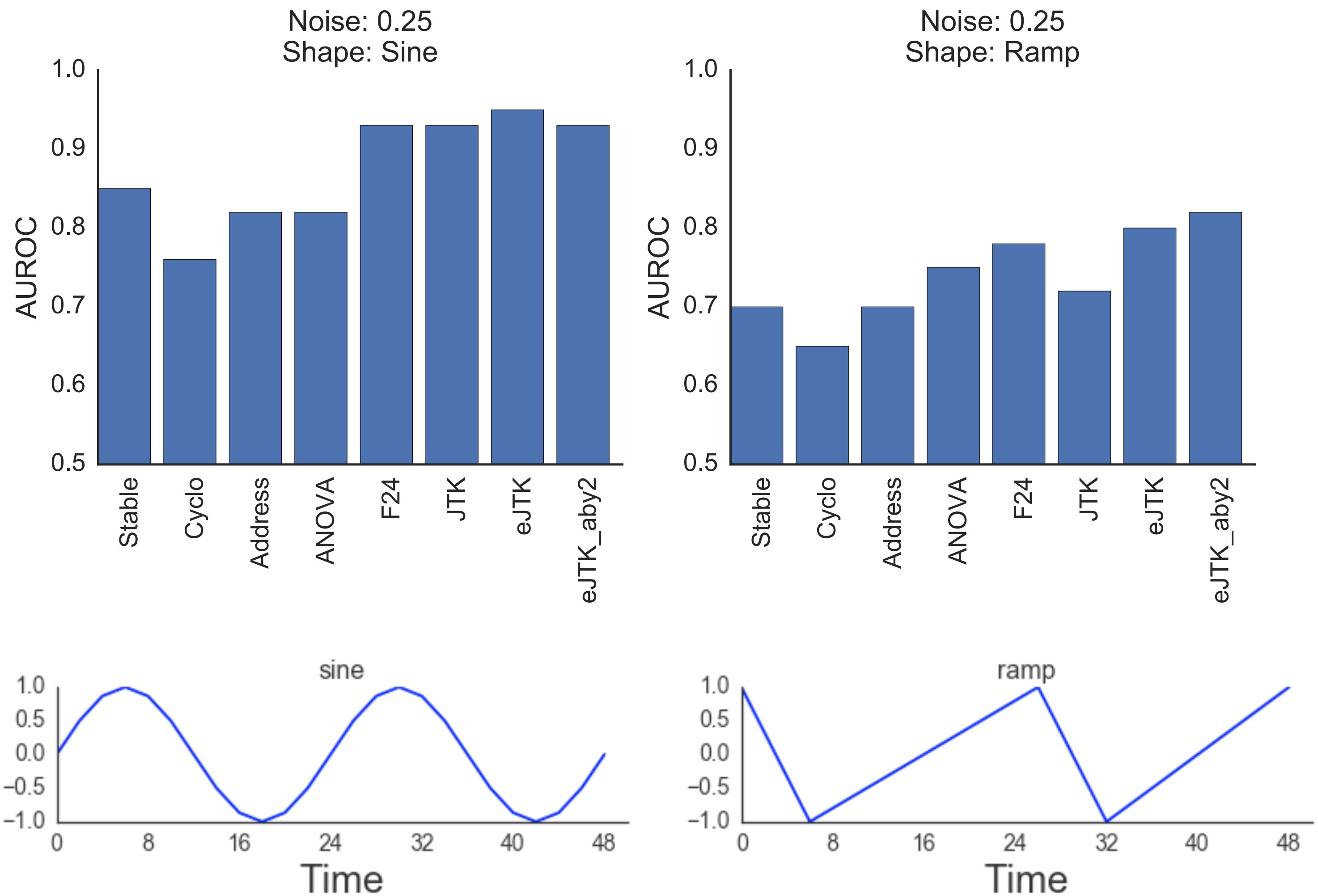




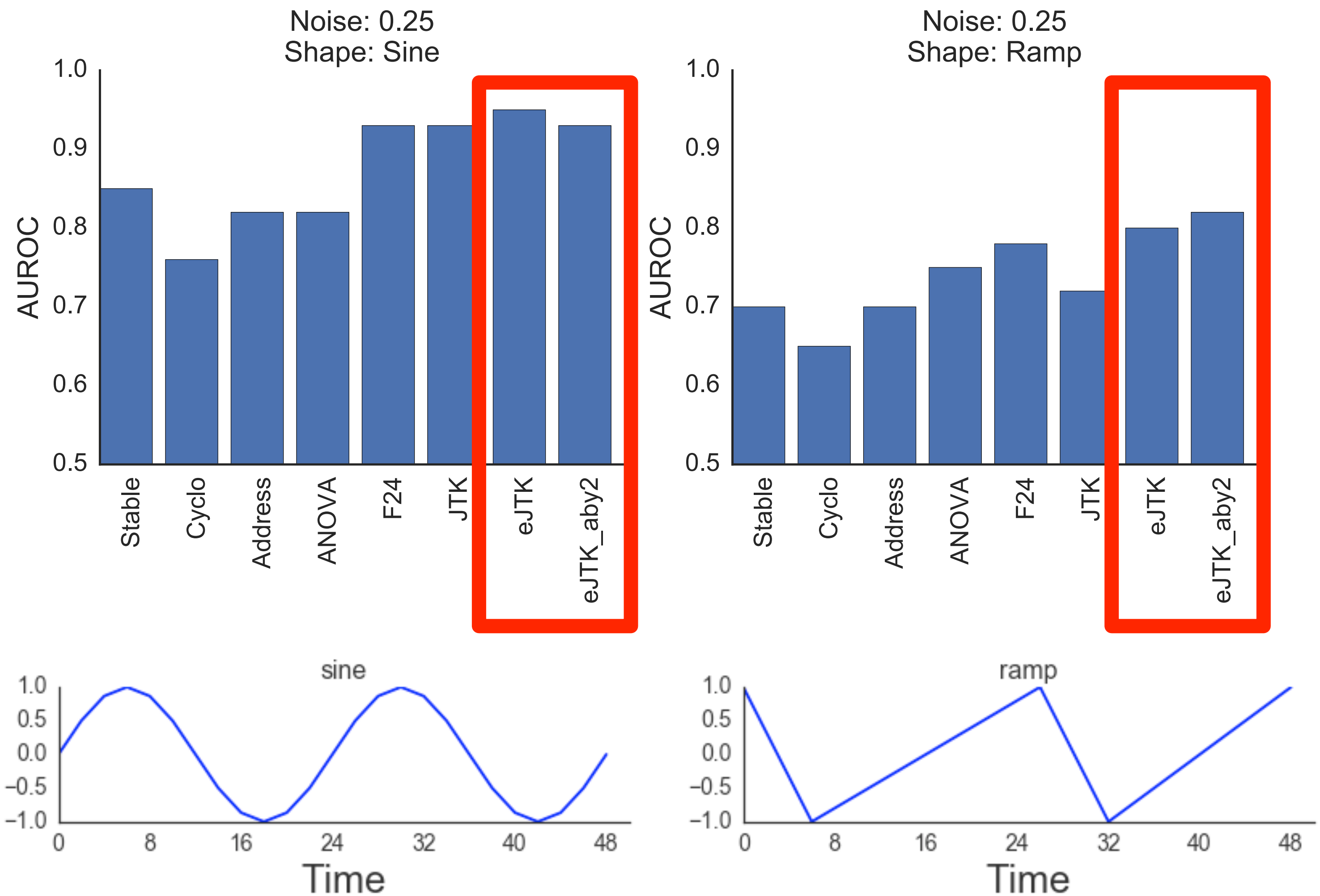




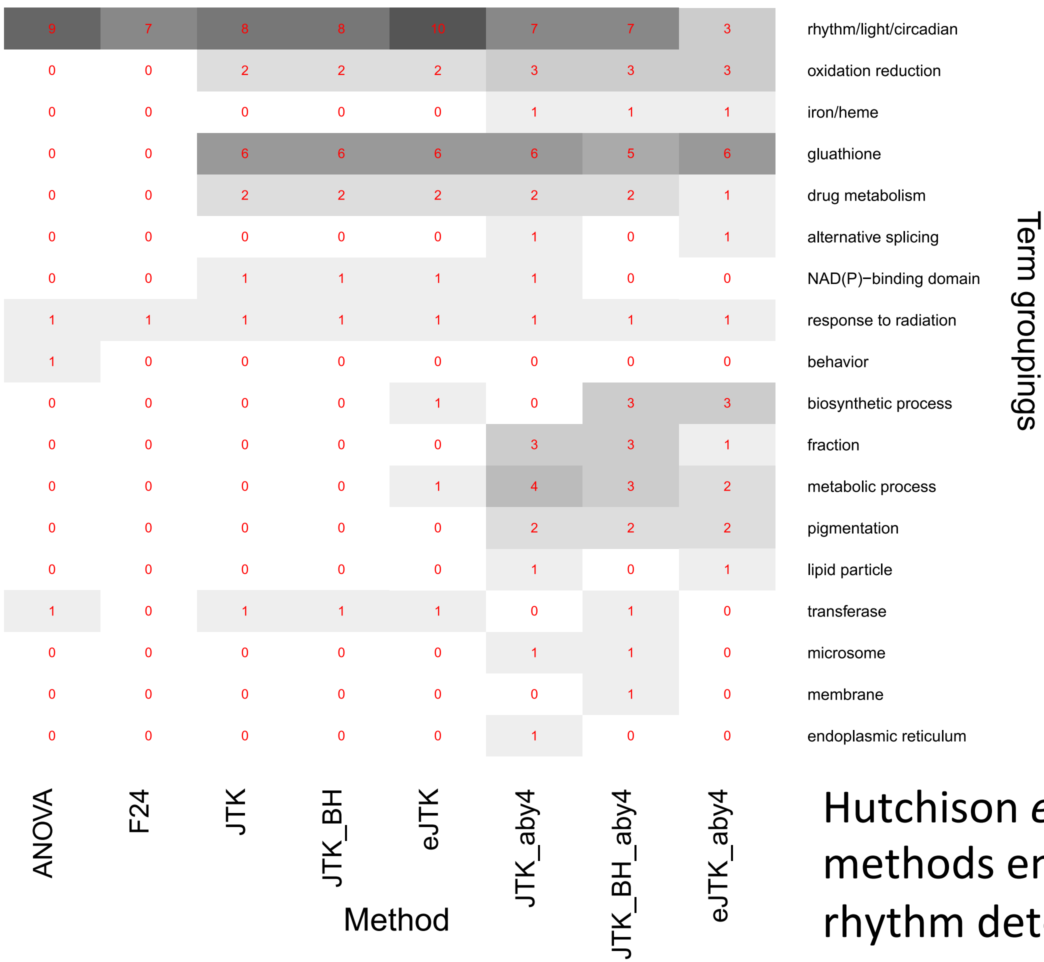
eJTK outperforms other methods on simulated data



eJTK outperforms other methods on simulated data



eJTK identifies ontologies missed by other methods



Hutchison *et al.* "Improved statistical methods enable greater sensitivity for rhythm detection in genome-wide data". *PLoS Computational Biology*. 2015 (11) 3

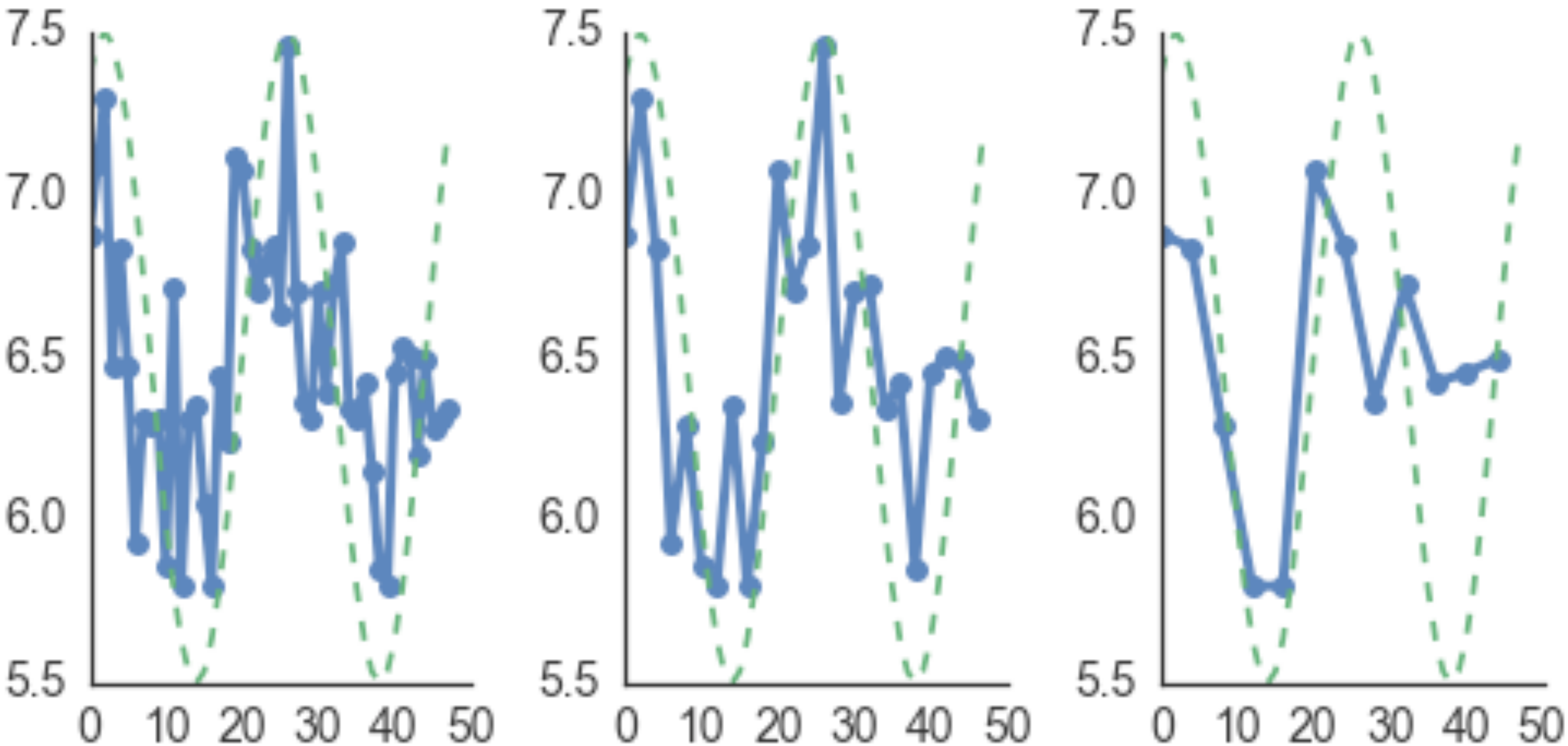
Outline

- Biological and Statistical Background
- Improvements to JTK_CYCLE
  - **Empirical JTK_CYCLE (eJTK)**
    - **Searching for asymmetric waveforms**
    - **Calculating accurate p-values**
    - **Hutchison *et al.* (2015) "Improved statistical methods enable greater sensitivity for rhythm detection in genome-wide data". *PLoS Computational Biology*. (11) 3**

Outline

- Biological and Statistical Background
- Improvements to JTK_CYCLE
  - Empirical JTK_CYCLE (eJTK)
  - **Bootstrap eJTK (BooteJTK)**
    - **Bootstrap resampling time series**
    - **Empirical Bayes variance estimation**

Three challenges of rhythm detection



- Sparse sampling of data
- High noise of measurements
- High rate of arrhythmic genes

Bootstrap resampling to propagate uncertainty from expression to rhythmicity

Initial time series data

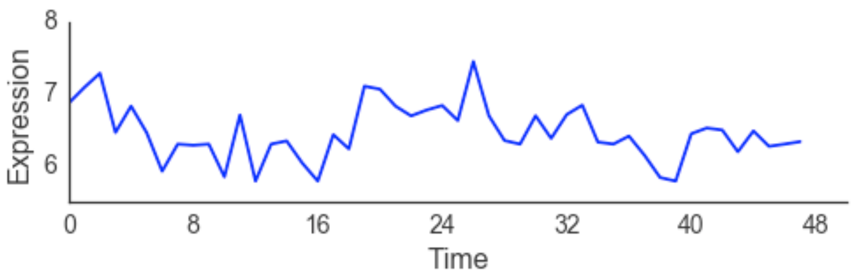
↓

Uncertainty in expression measurement

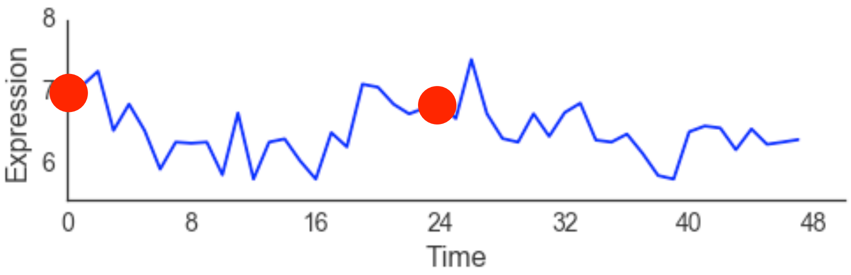
↓

Uncertainty in rhythmicity

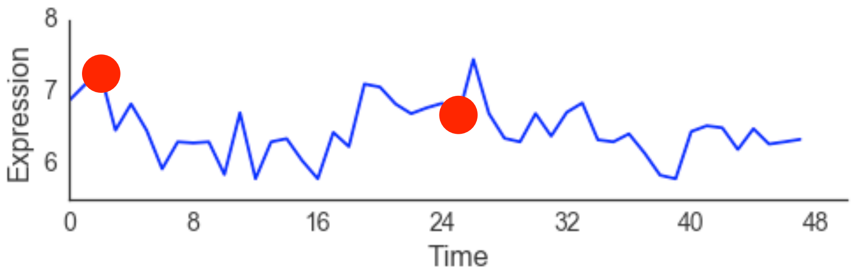
Averaging data to get error bars



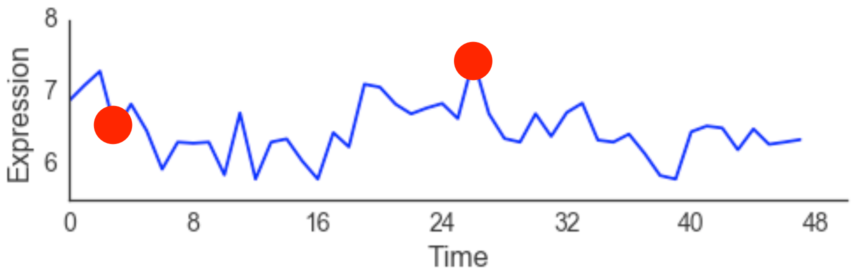
Averaging data to get error bars



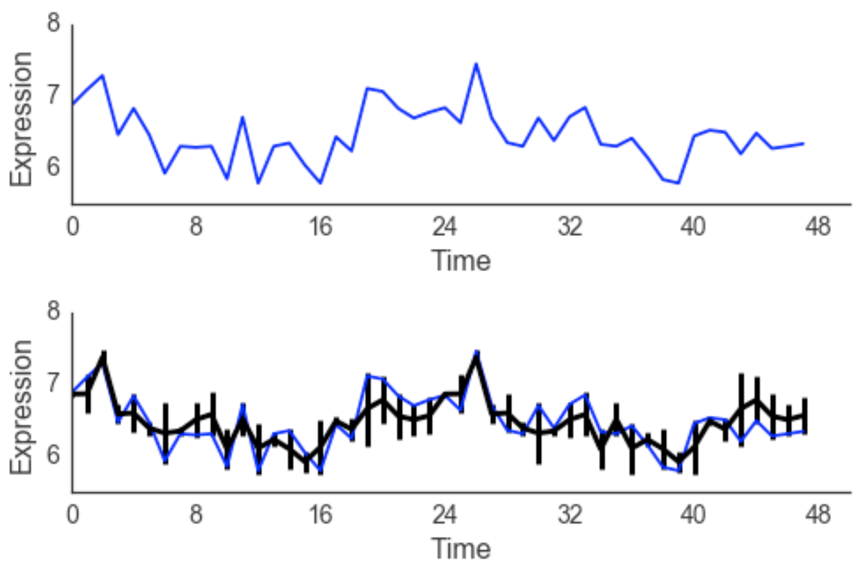
Averaging data to get error bars



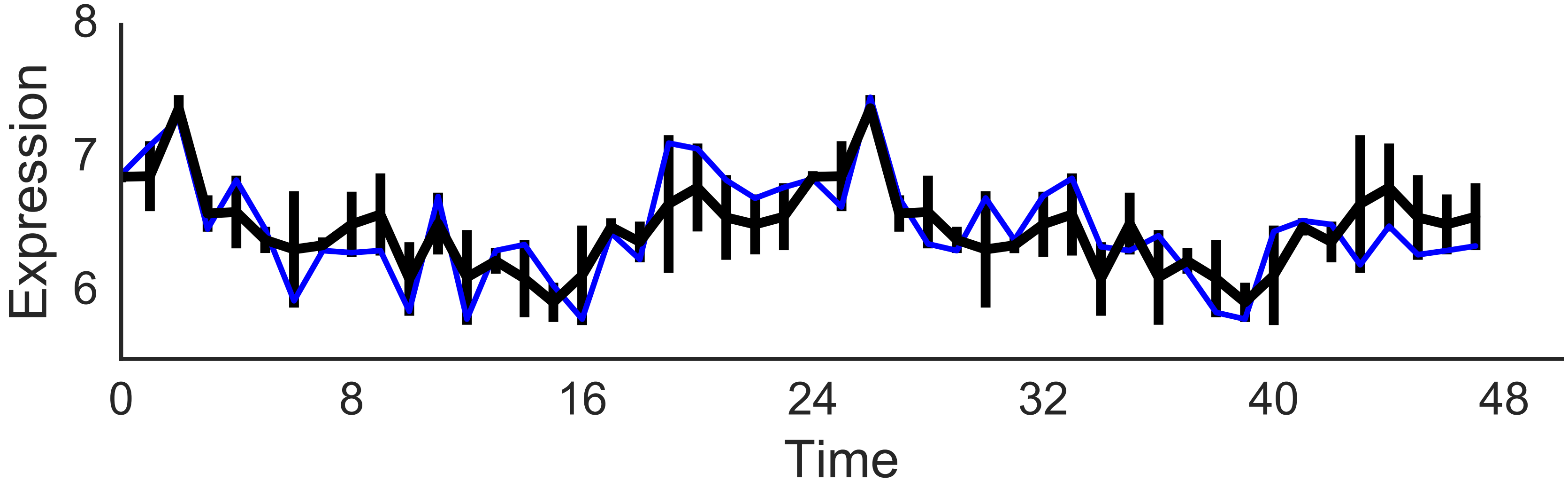
Averaging data to get error bars



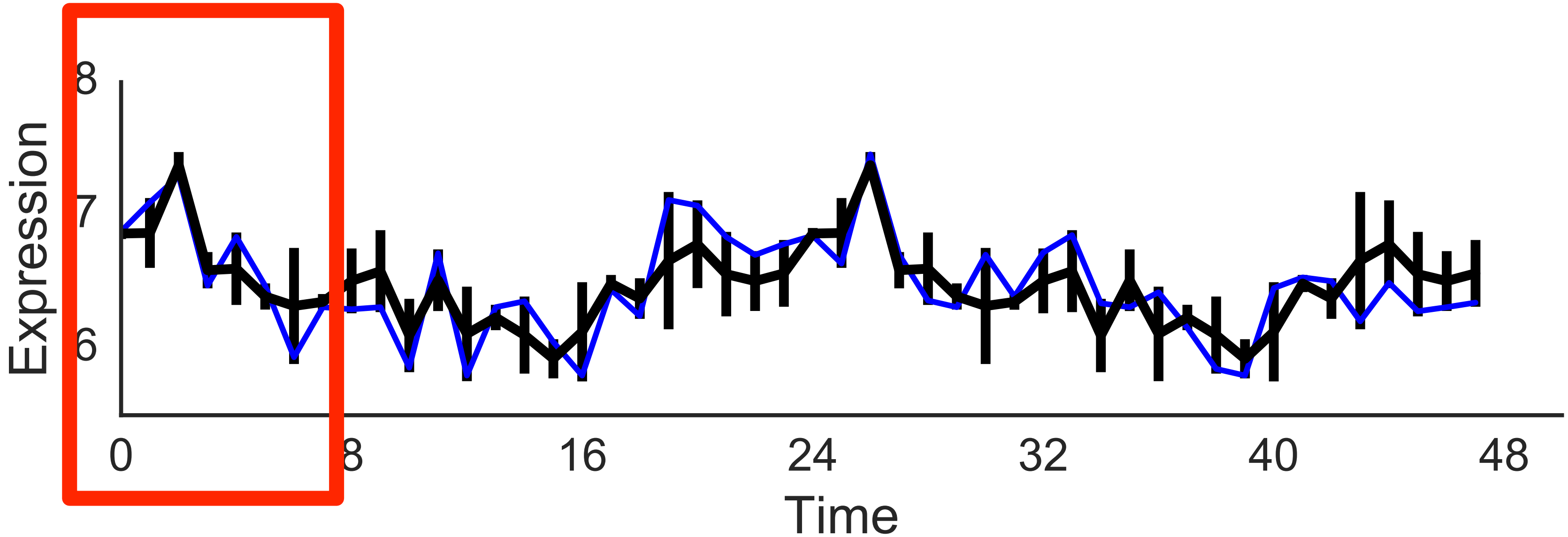
Averaging data to get error bars



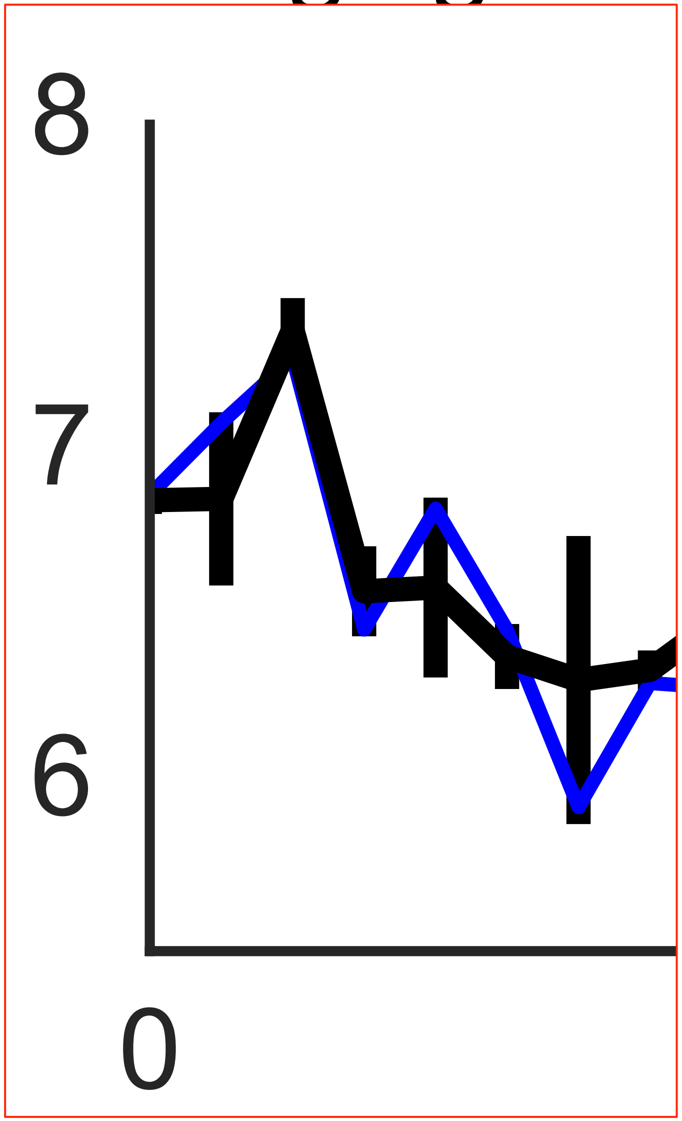
Averaging data to get error bars



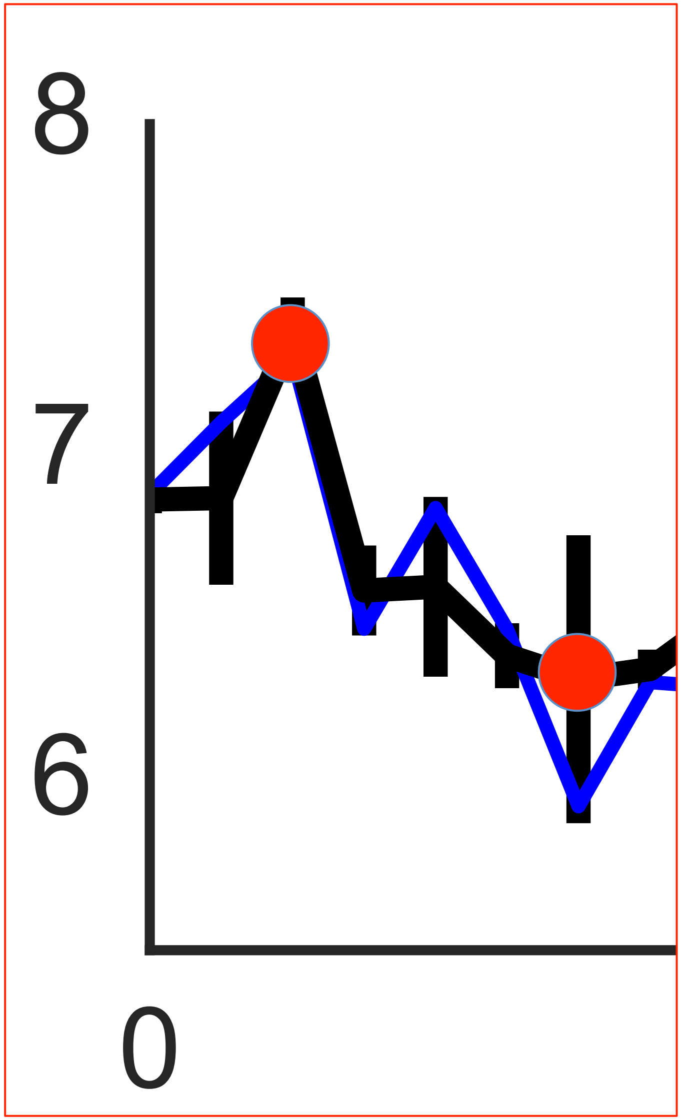
Averaging data to get error bars



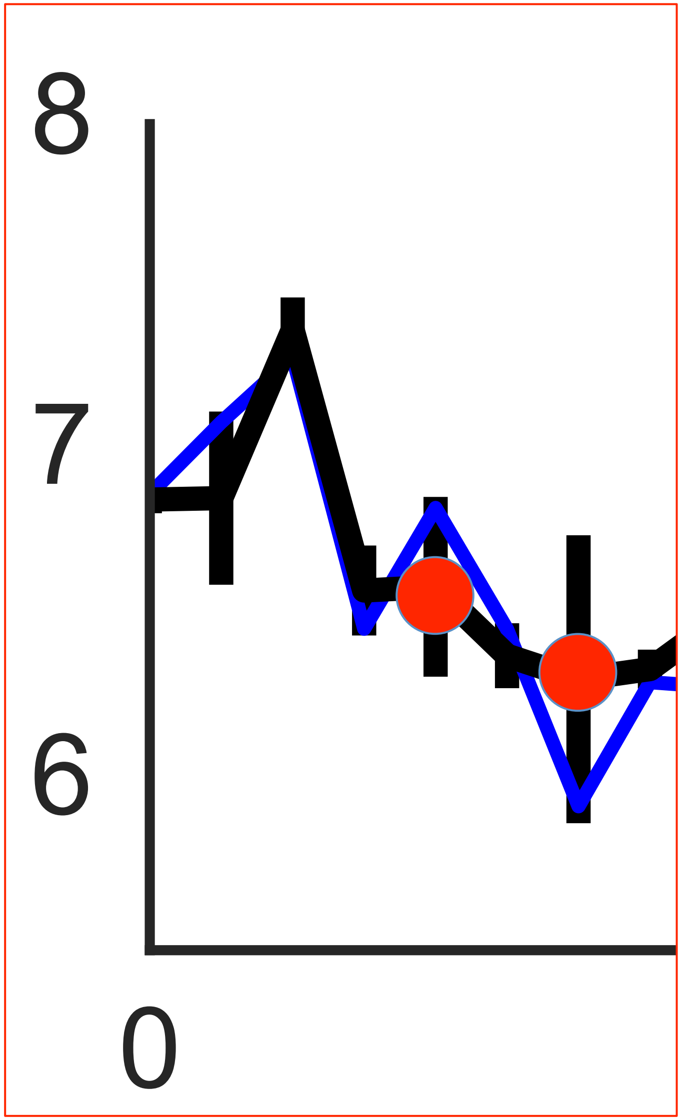
Averaging data to get error bars



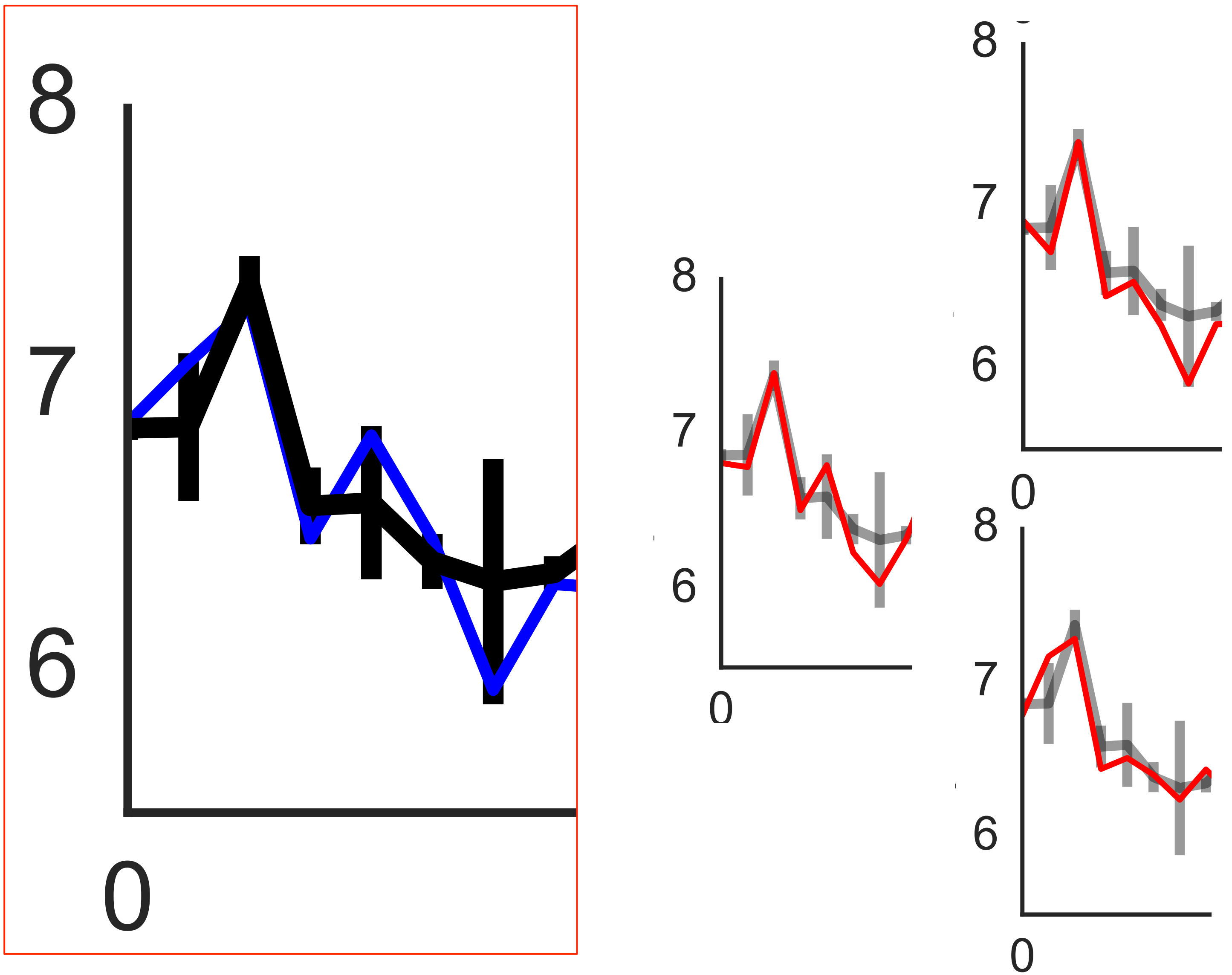
Can we measure uncertainty in ordering?



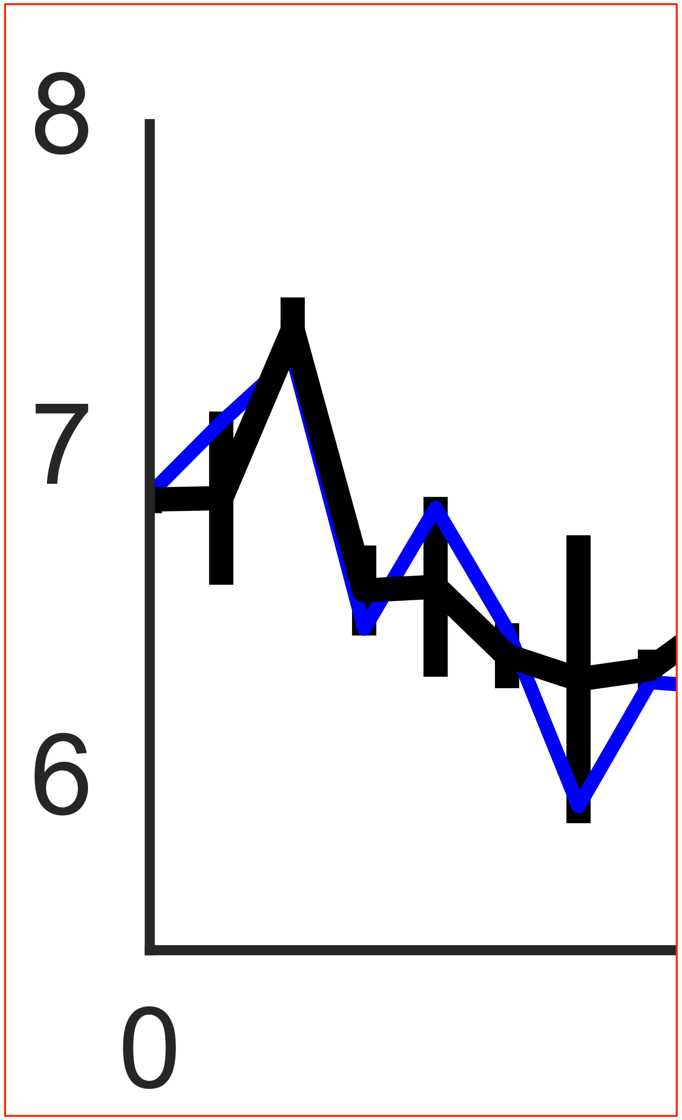
Can we measure uncertainty in ordering?

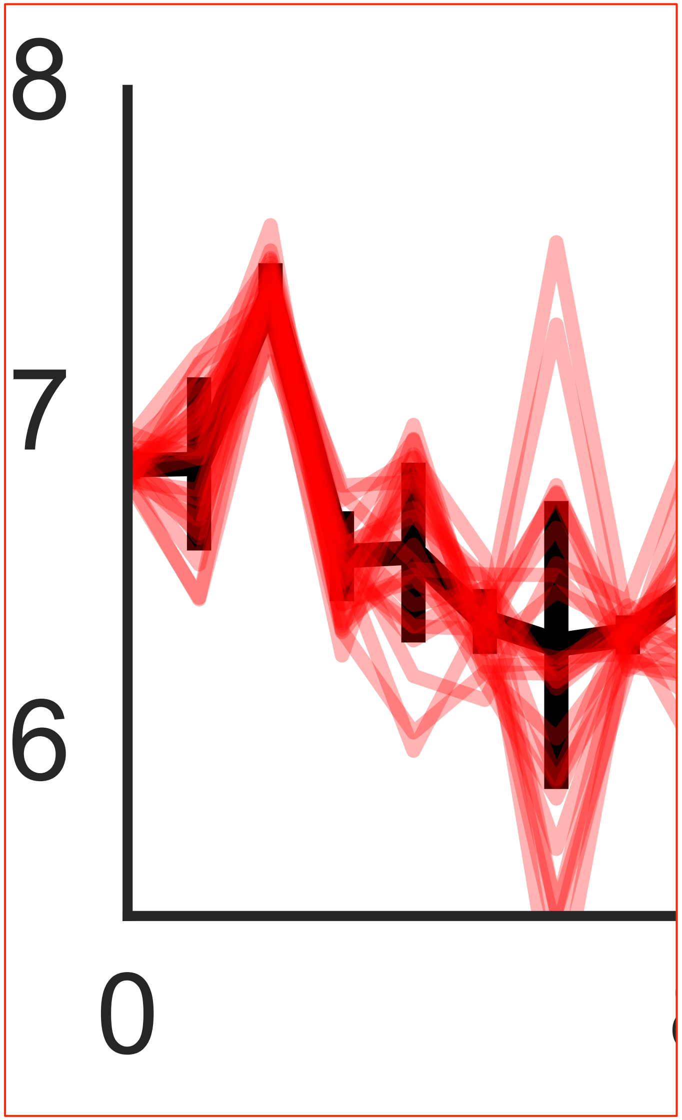


Resample time series to 'replicate' experiment

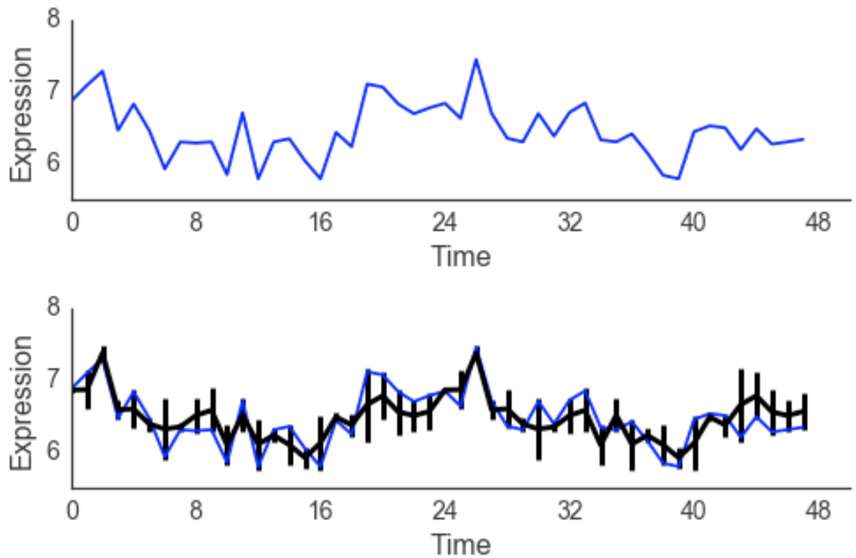


Resample time series to 'replicate' experiment

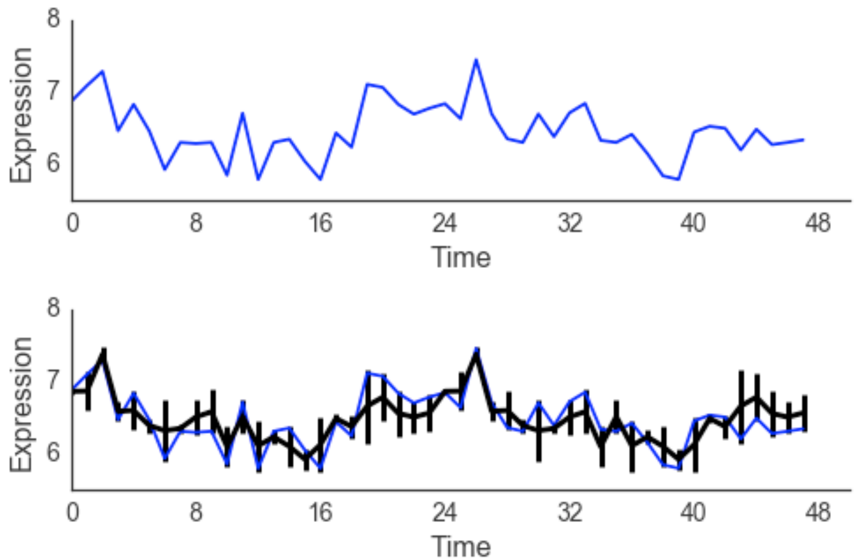


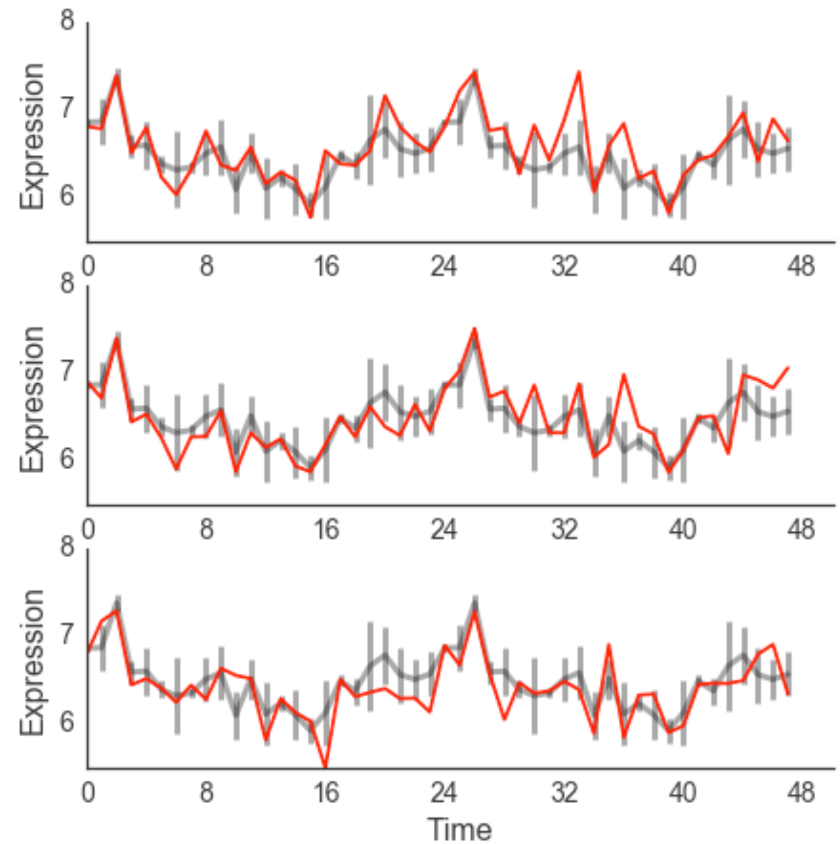


Resample time series to ‘replicate’ experiment

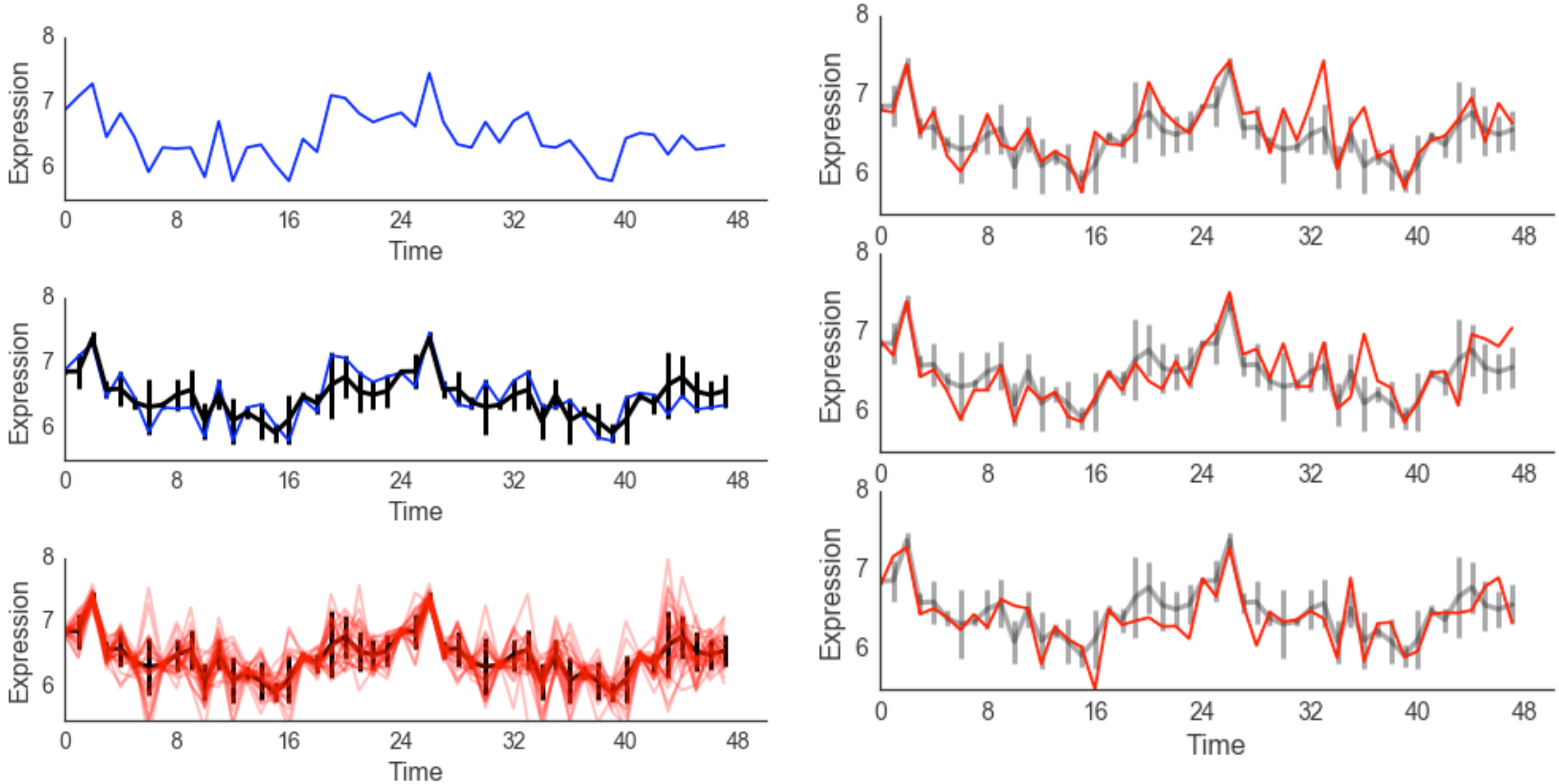


Resample from each point to obtain simulated time series

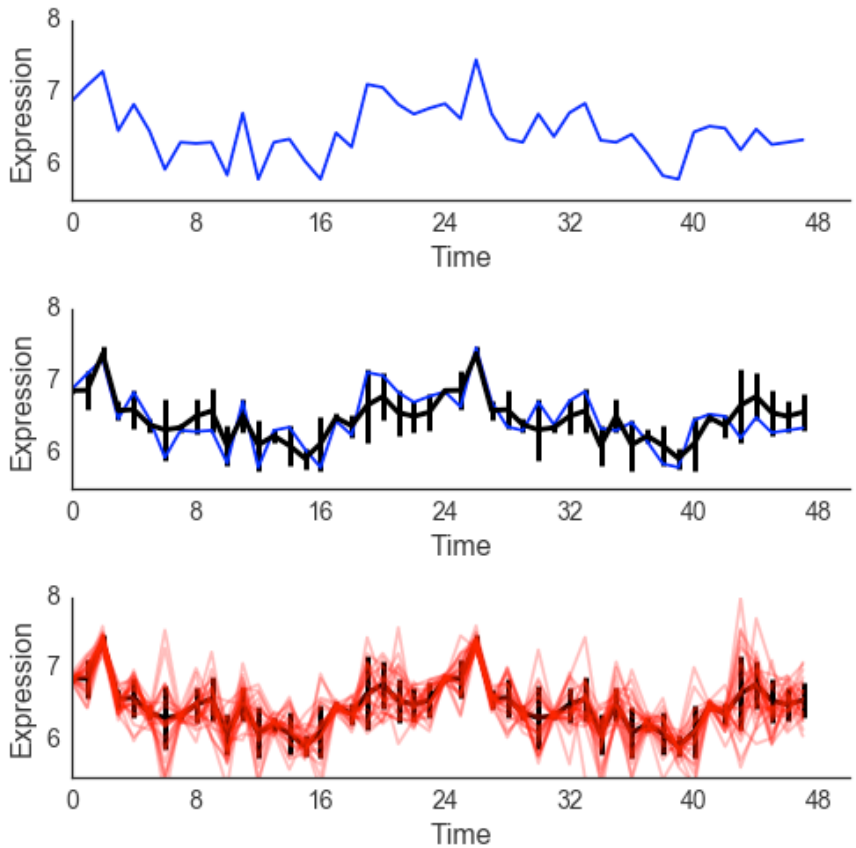


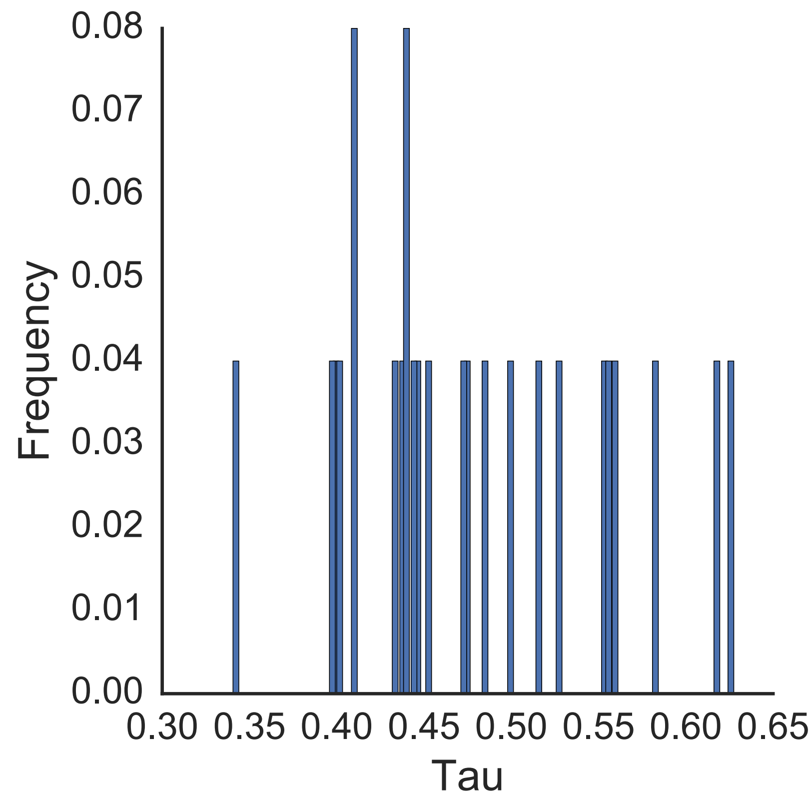


Resample from each point to obtain simulated time series

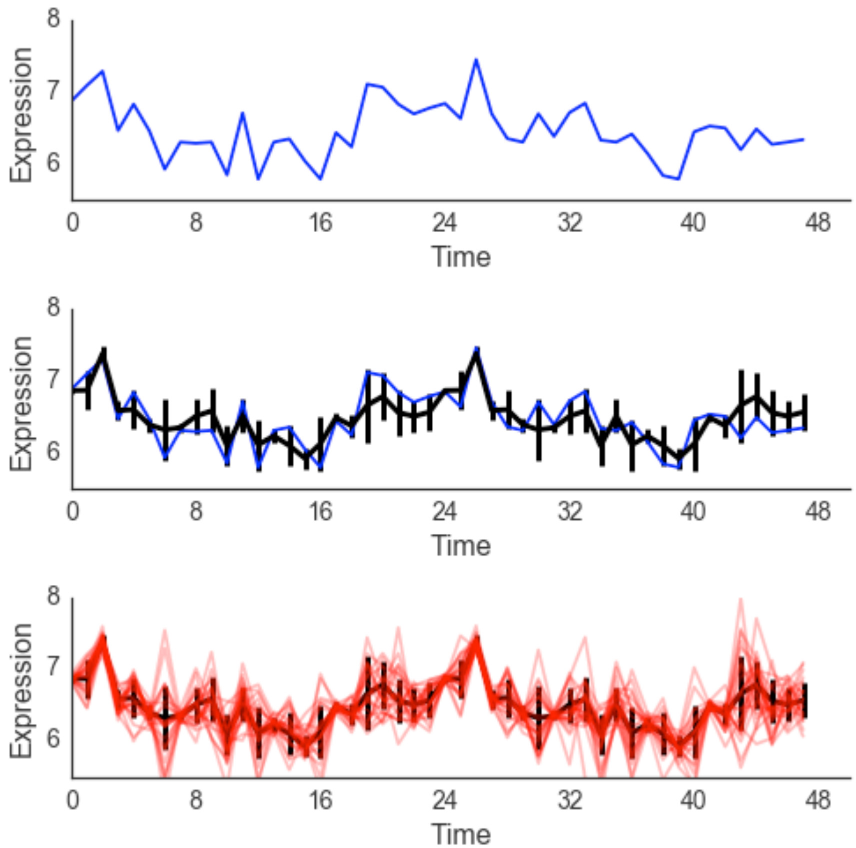


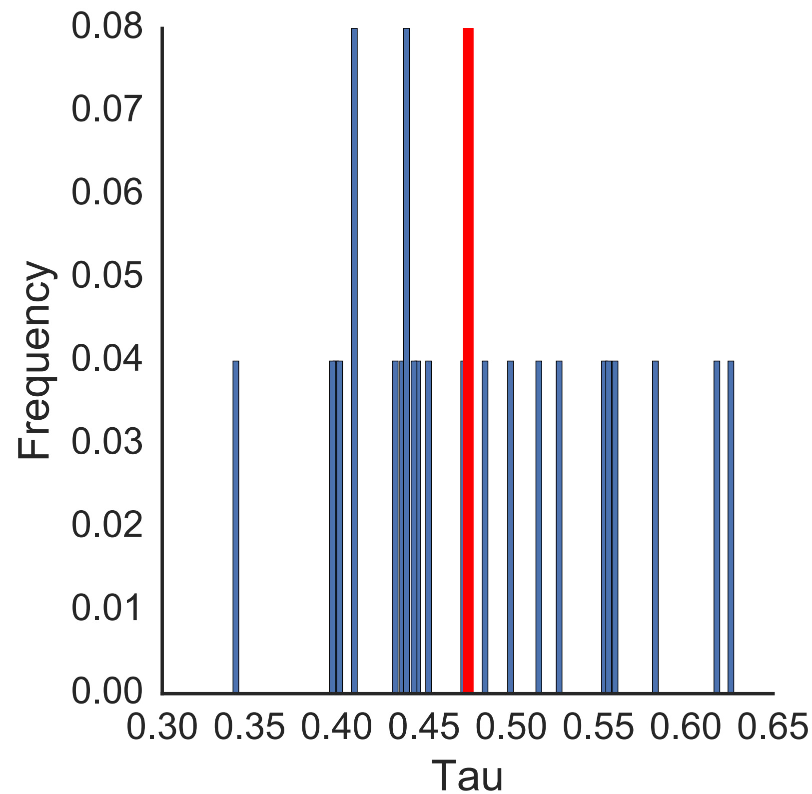
Run eJTK on each resampled time series to get distribution of Tau values





Average Tau values to get summary statistic mean value

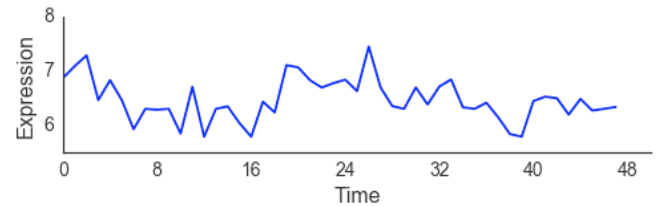




Bootstrap resampling to propagate uncertainty from expression to rhythmicity

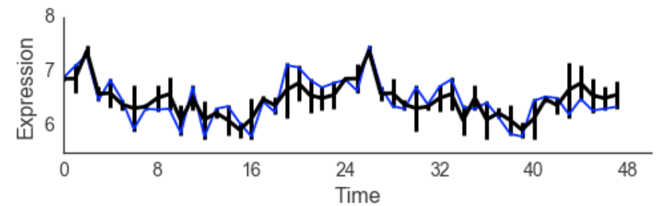
Initial time series data

Average data



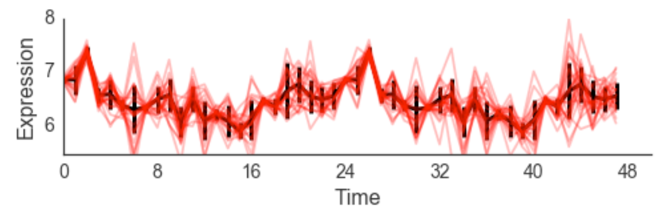
Uncertainty in expression measurement

Bootstrap replicates



Uncertainty in expression ordering

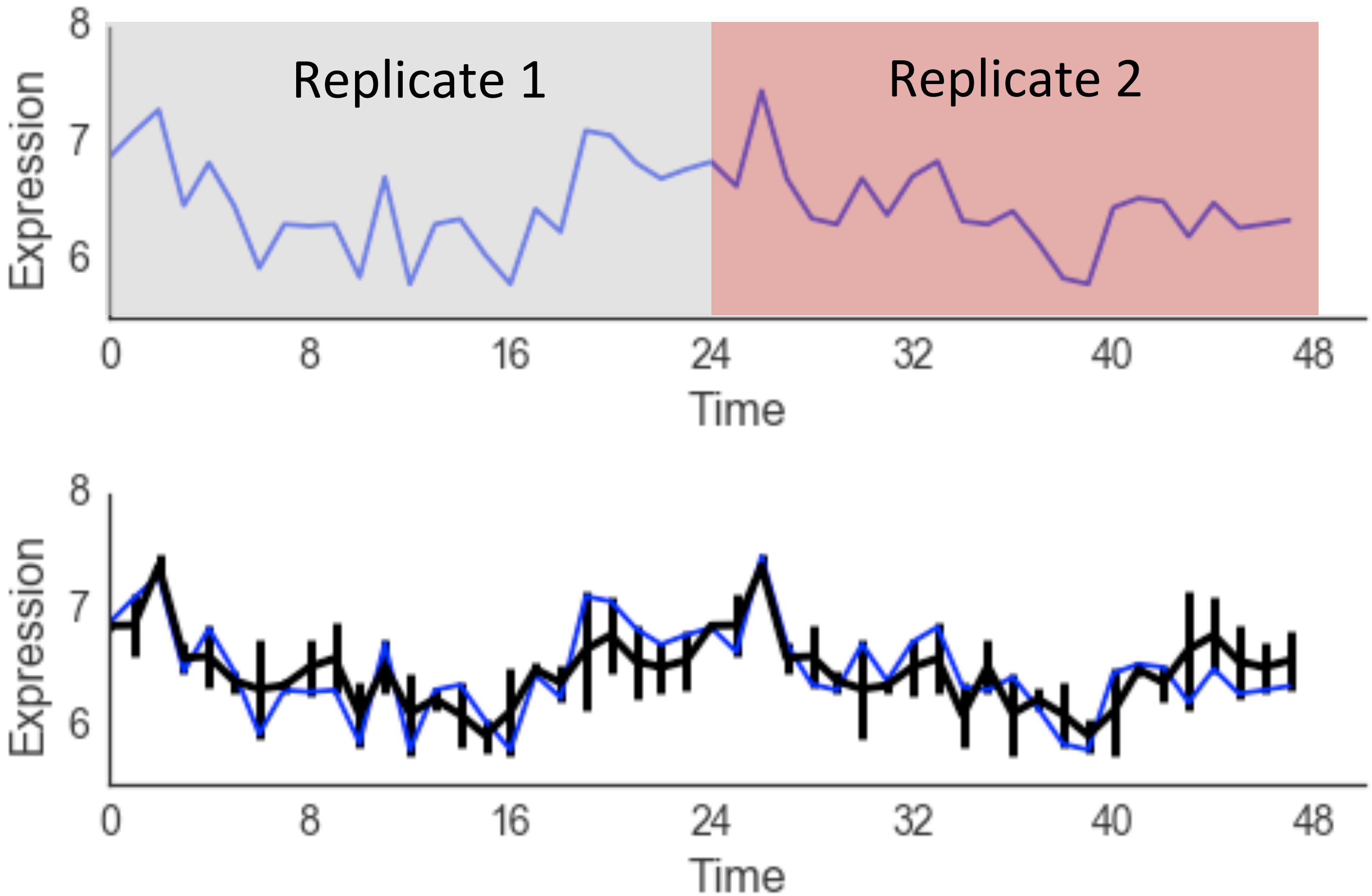
Run eJTK on replicates



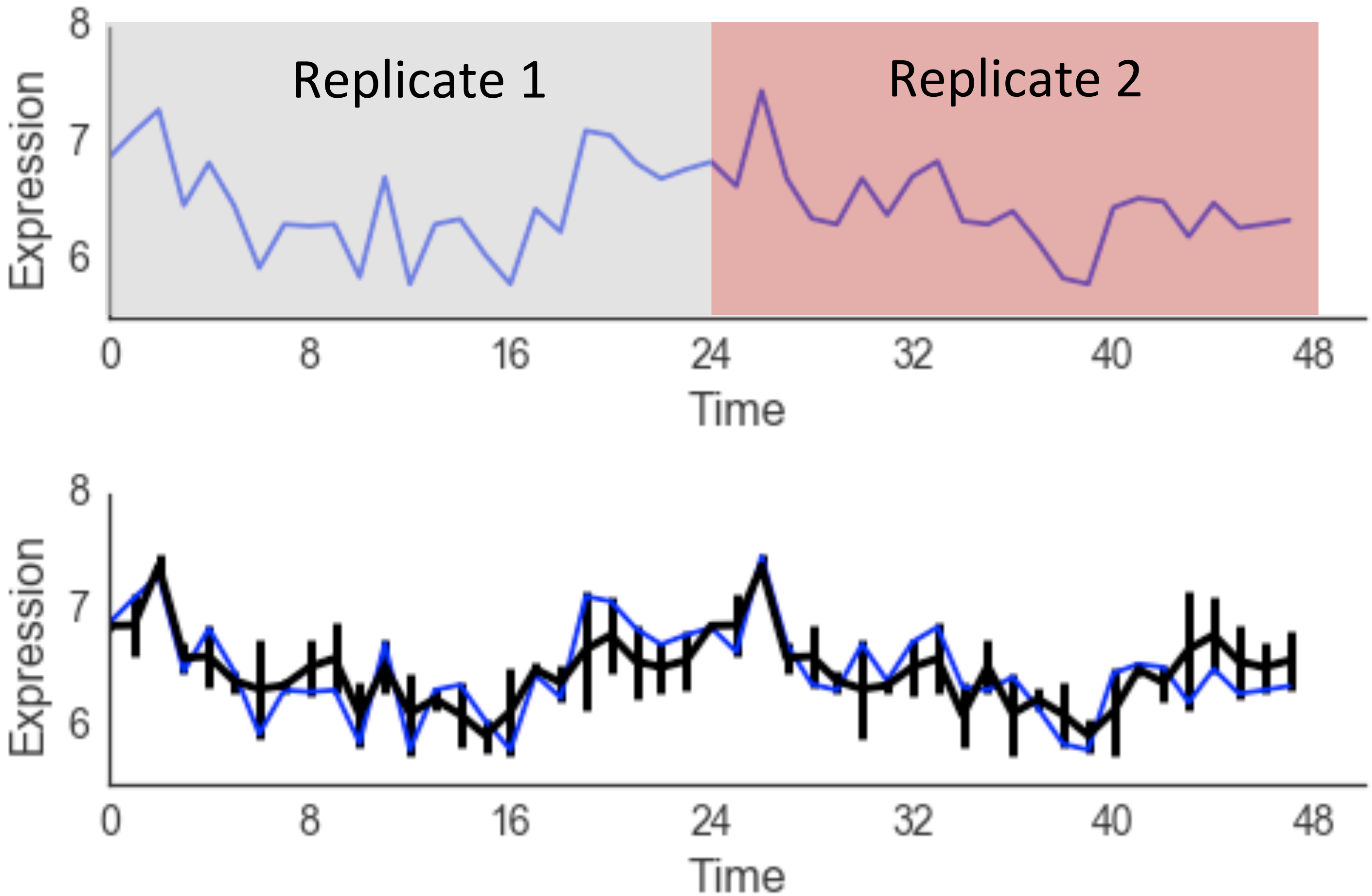
Uncertainty in rhythmicity → Summary statistic

Average distribution

Low replicate numbers reduce confidence in variance estimates



Empirical Bayes approach: Improving variance estimates by pooling them



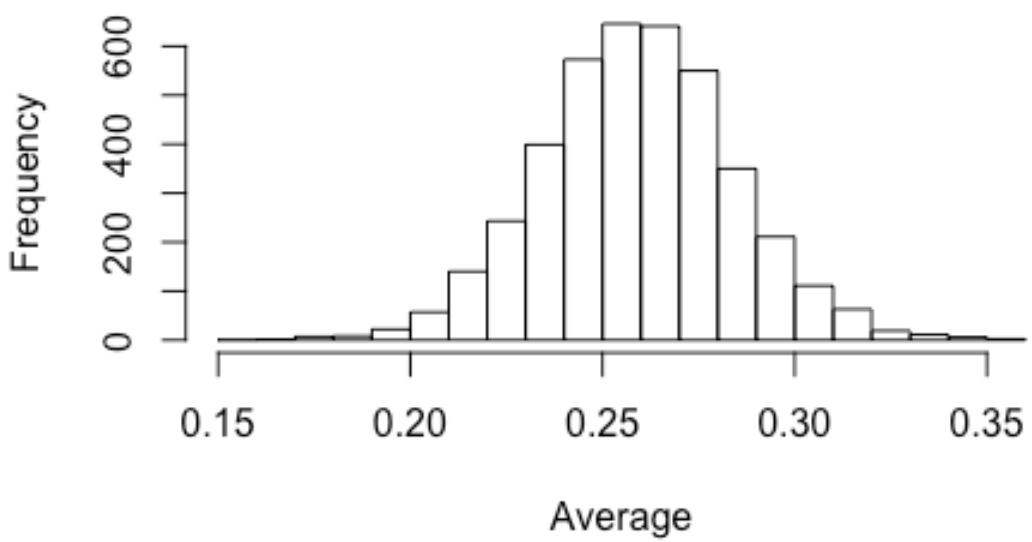
Empirical Bayes: Baseball batting average analogy

Player	Hits	At-Bats	Avg.
A	4	10	0.400
B	30	100	0.300
C	250	1000	0.250

Modeled after “Understanding empirical Bayes estimation (using baseball statistics)”
David Robinson Sept 30, 2015

Empirical Bayes: Baseball batting average analogy

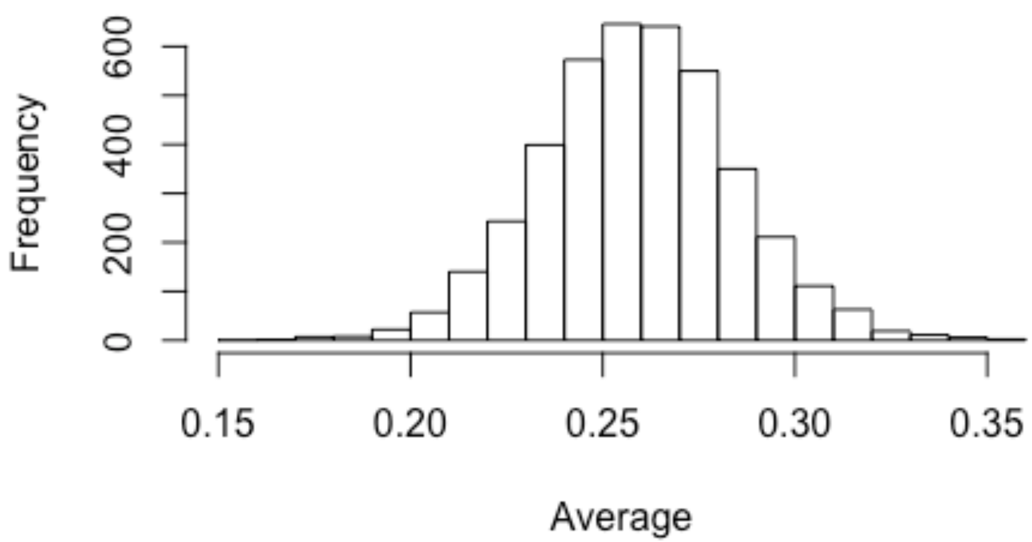
Player	Hits	At-Bats	Avg.
A	4	10	0.400
B	30	100	0.300
C	250	1000	0.250



Modeled after “Understanding empirical Bayes estimation (using baseball statistics)” David Robinson Sept 30, 2015

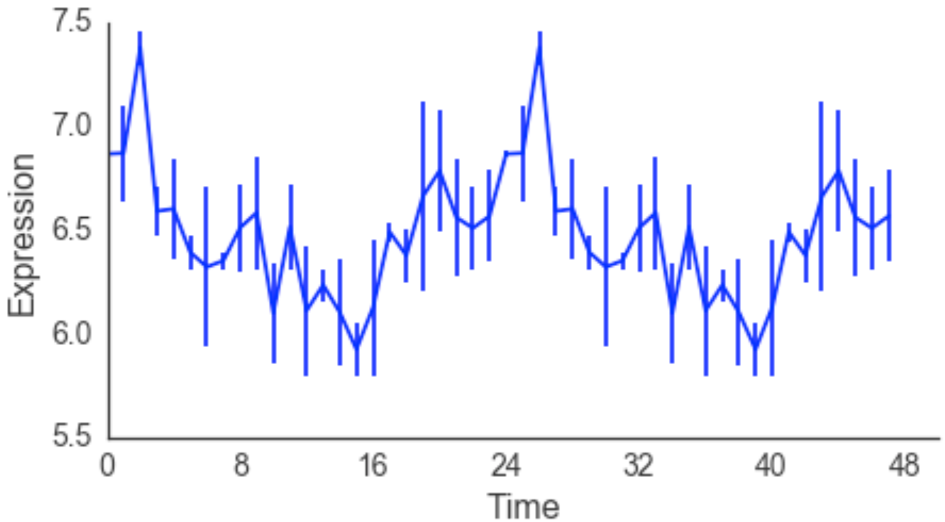
Empirical Bayes: Baseball batting average analogy

Player	Hits	At-Bats	Avg.	Adj. Avg.
A	4	10	0.400	0.263
B	30	100	0.300	0.269
C	250	1000	0.250	0.252

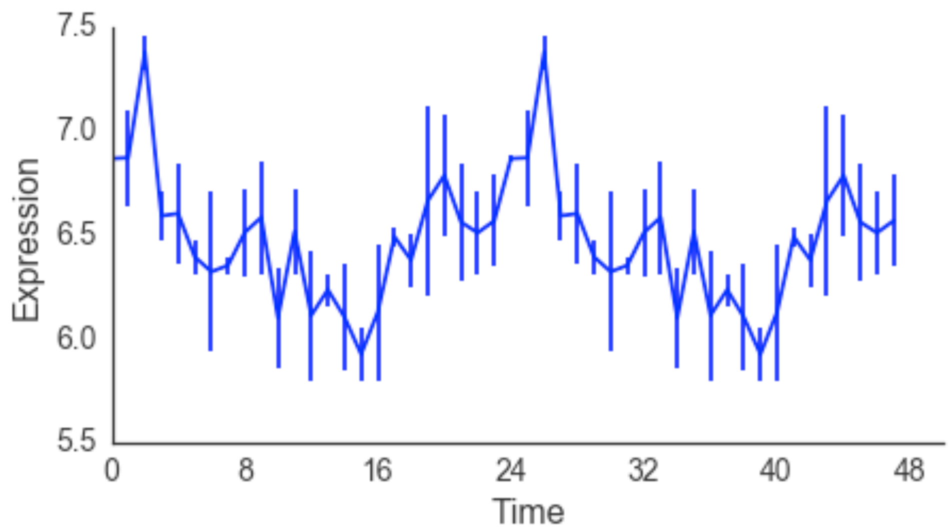


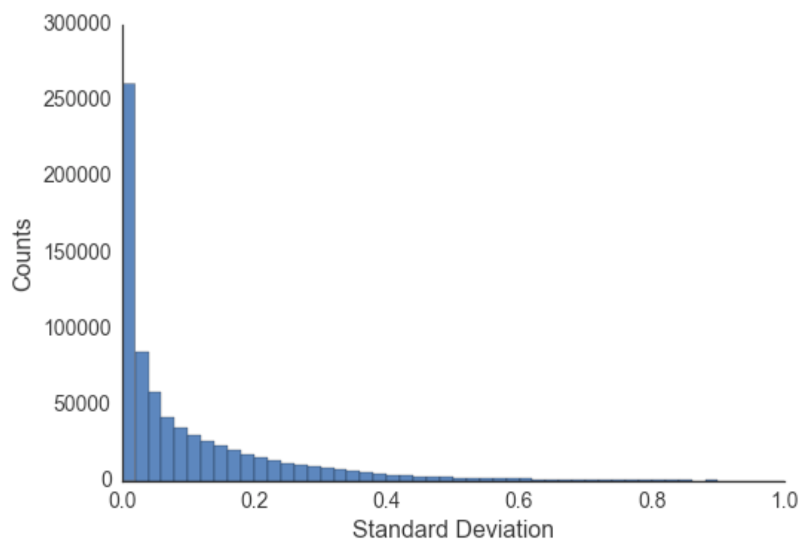
Modeled after "Understanding empirical Bayes estimation (using baseball statistics)"
David Robinson Sept 30, 2015

Empirical Bayes approach: Improving variance estimates by pooling them

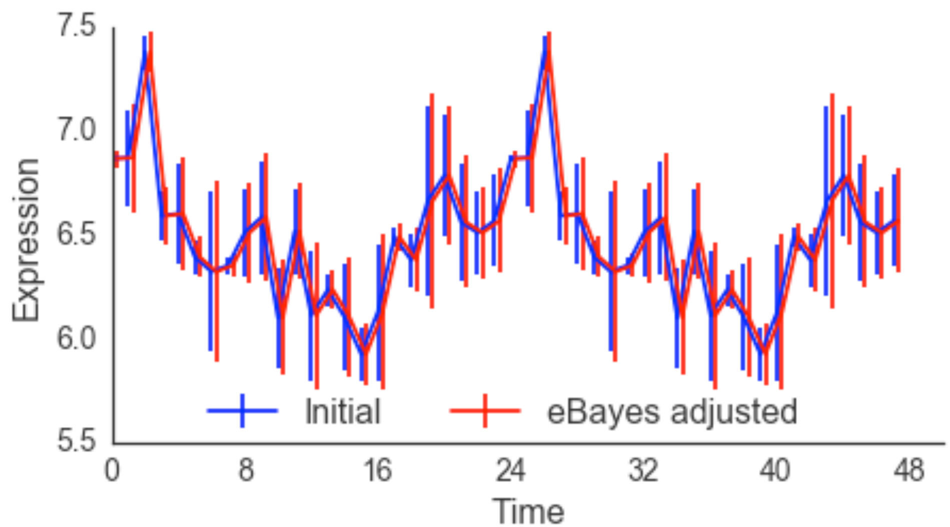


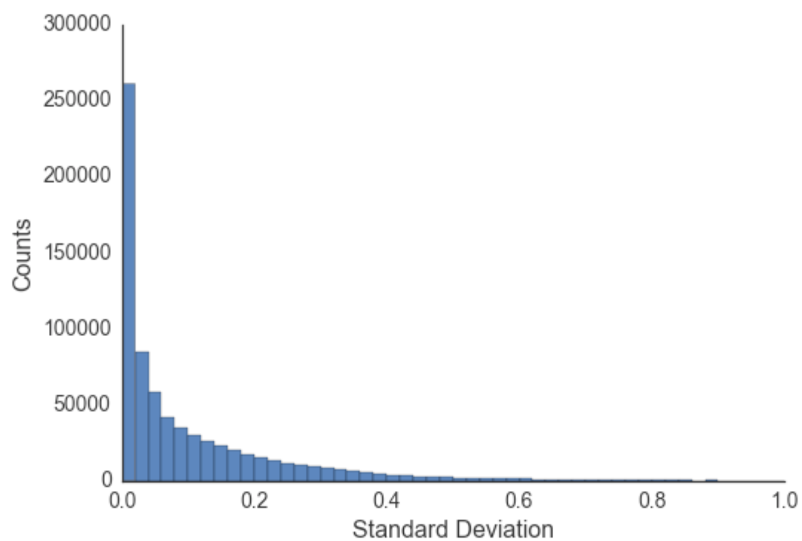
Empirical Bayes approach: Improving variance estimates by pooling them





Empirical Bayes approach: Improving variance estimates by pooling them



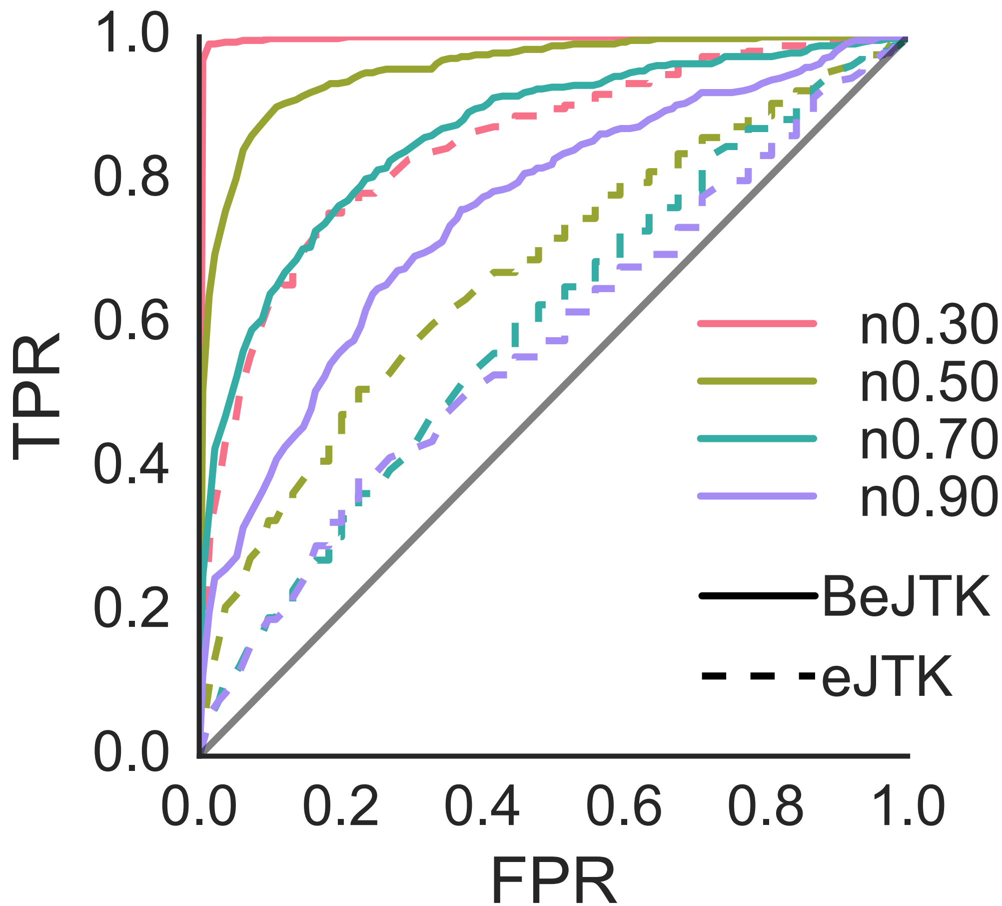


Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.

Outline

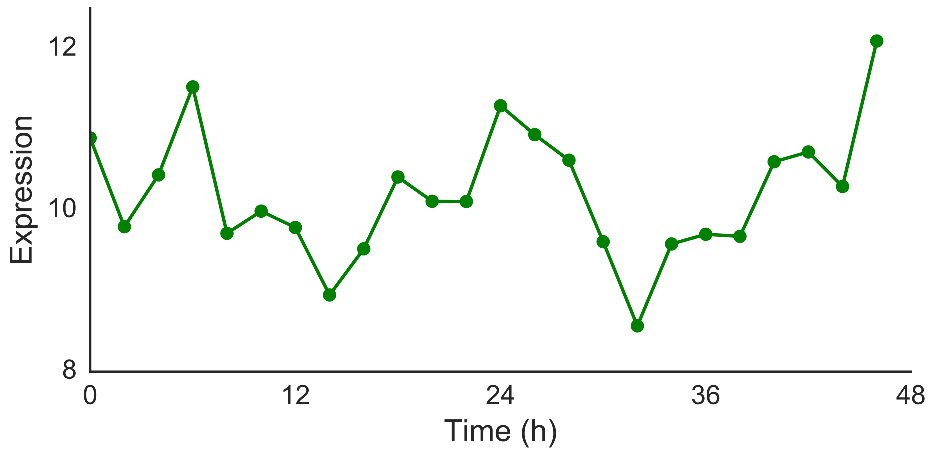
- Biological and Statistical Background
- Improvements to JTK_CYCLE
  - Empirical JTK_CYCLE (eJTK)
  - **Bootstrap eJTK (BooteJTK)**
    - **Bootstrap resampling time series**
      - **New to rhythm detection**
    - **Empirical Bayes variance estimation**
      - **Common in differential expression analysis**
      - **New to rhythm detection**

BooteJTK outperforms eJTK on simulated data

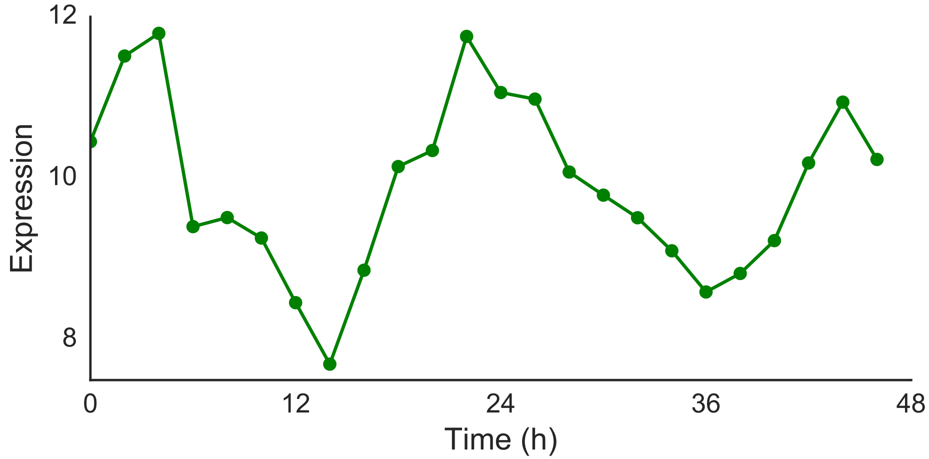


BooteJTK is sensitive to the intra-point variance relative to the variance of the entire time series

eJTK Tau: 0.57
BooteJTK Tau: 0.67

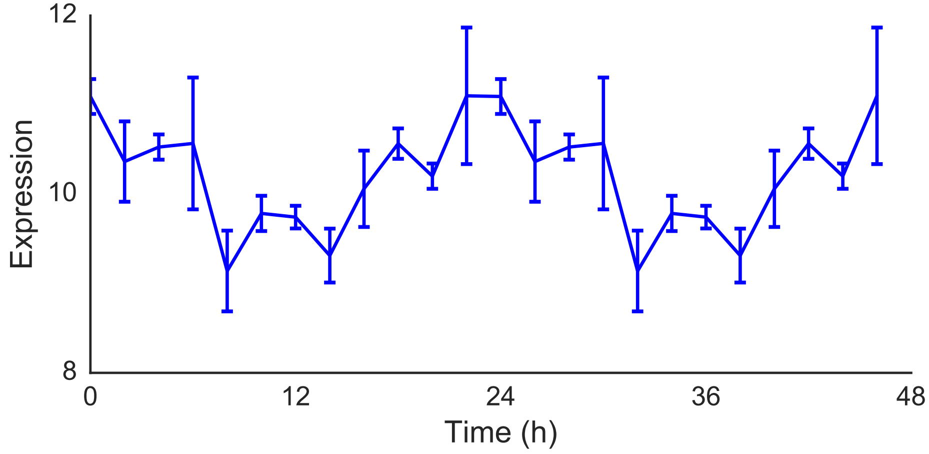


eJTK Tau: 0.57
BooteJTK Tau: 1.08

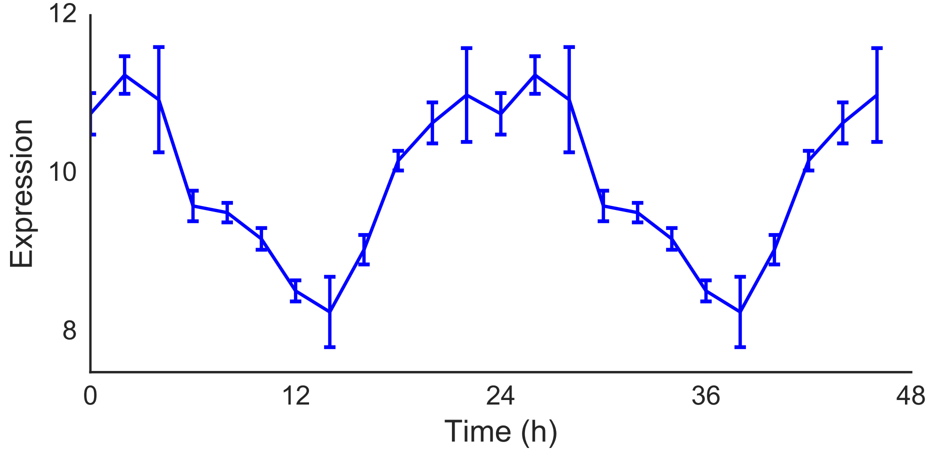


BooteJTK is sensitive to the intra-point variance relative to the variance of the entire time series

eJTK Tau: 0.57
BooteJTK Tau: 0.67

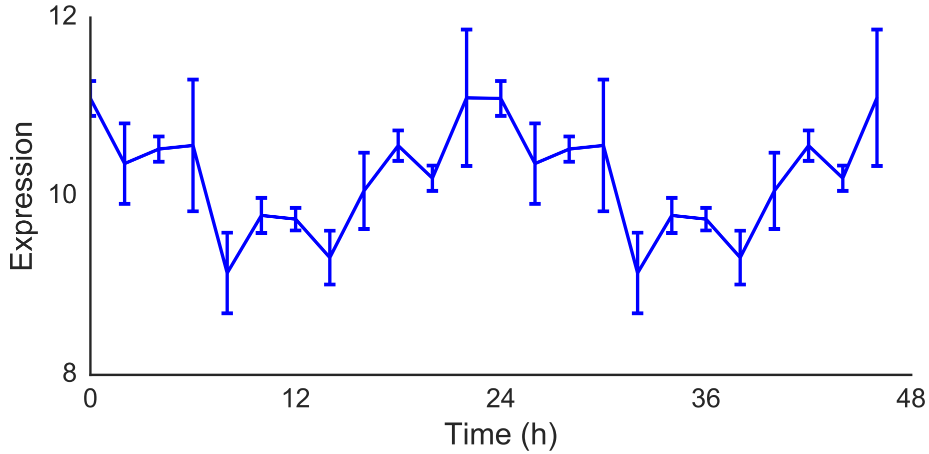


eJTK Tau: 0.57
BooteJTK Tau: 1.08

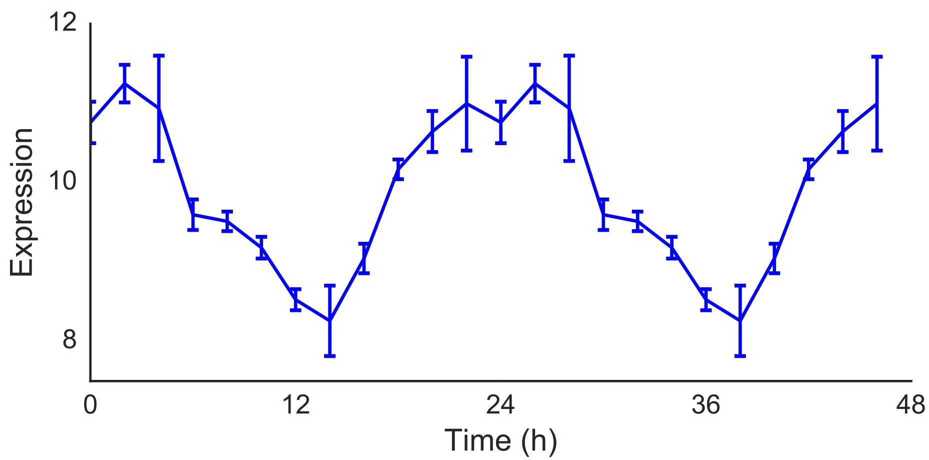


BooteJTK is sensitive to the intra-point variance relative to the variance of the entire time series

eJTK Tau: 0.57
BooteJTK Tau: 0.67

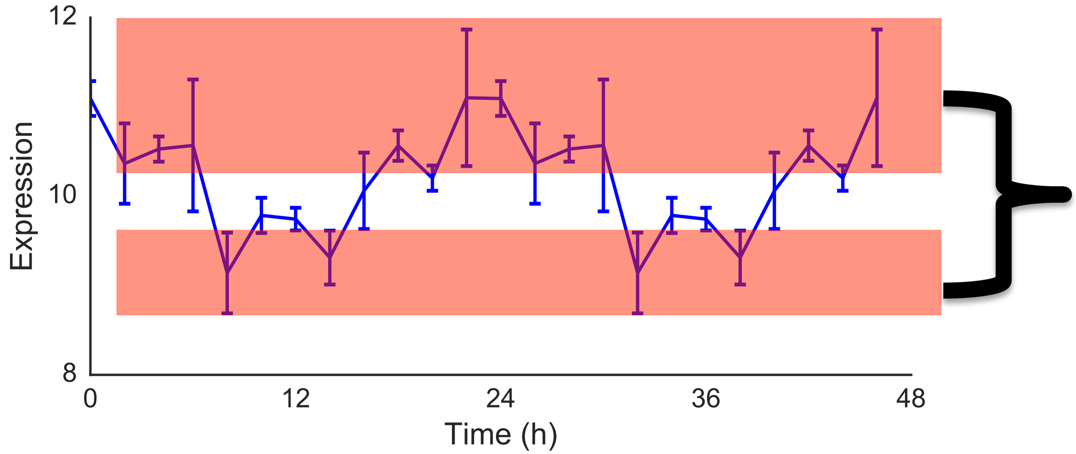


eJTK Tau: 0.57
BooteJTK Tau: 1.08

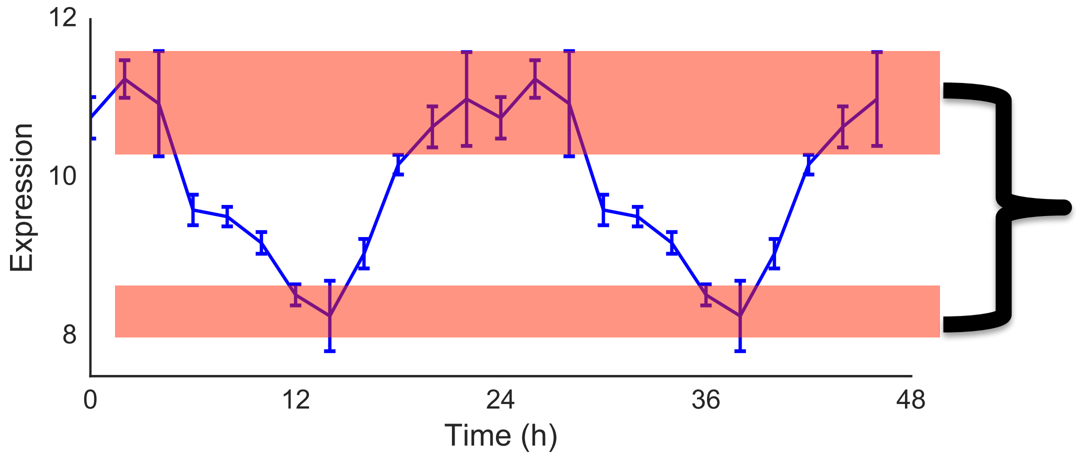


BooteJTK is sensitive to the intra-point variance relative to the variance of the entire time series

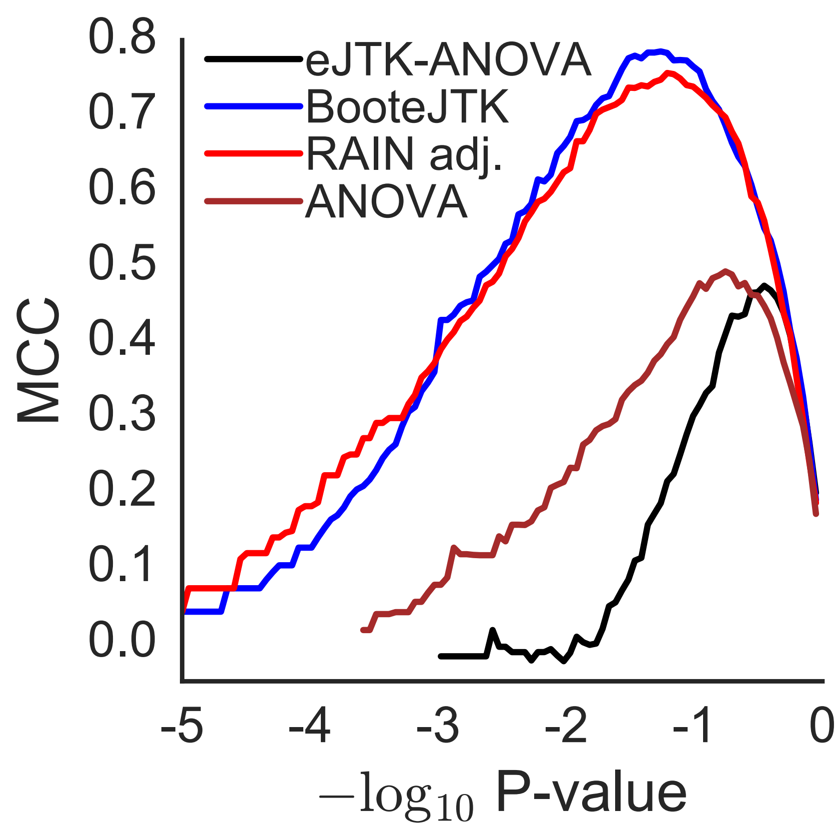
eJTK Tau: 0.57
BooteJTK Tau: 0.67



eJTK Tau: 0.57
BooteJTK Tau: 1.08

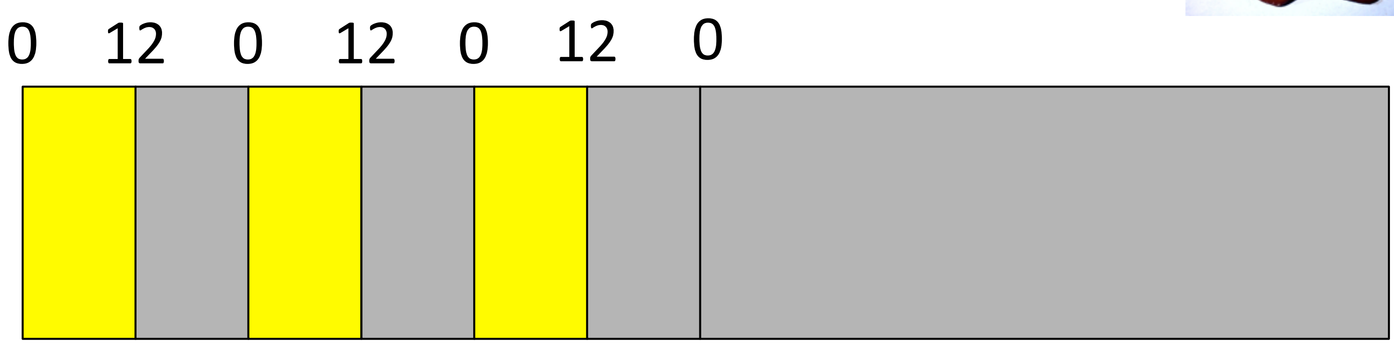


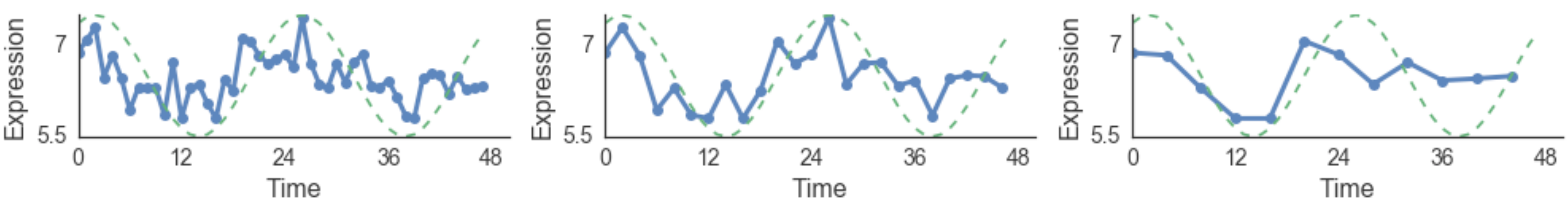
BooteJTK outperforms alternative methods, including a combination of eJTK and ANOVA



Hughes *et al.* 1h liver dataset

12 h light
12 h dark



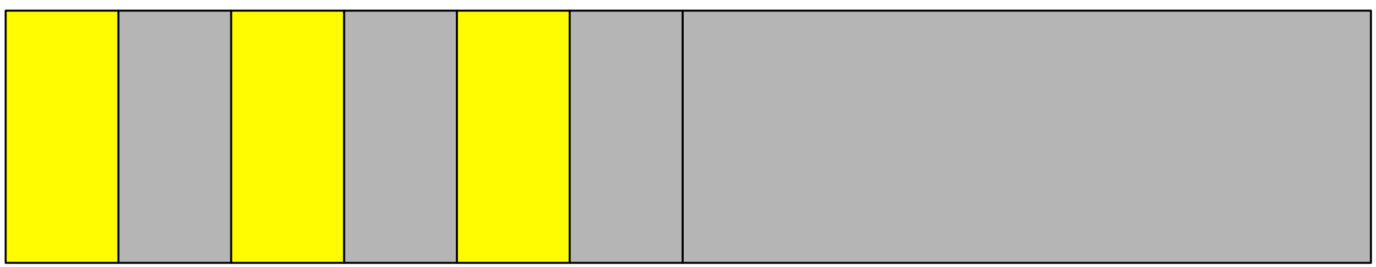


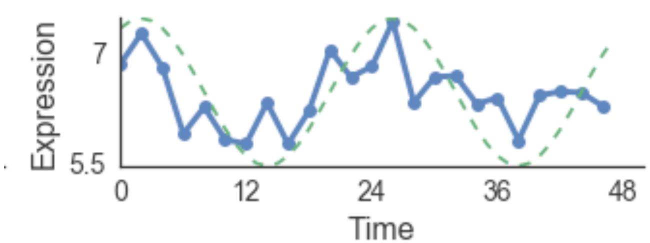
Hughes *et al.* (2009) “Harmonics of Circadian Gene Transcription in Mammals.”
PLoS Genetics, 2009. 5(4): e1000442

Hughes *et al.* 1h liver dataset

12 h light
12 h dark

0 12 0 12 0 12 0

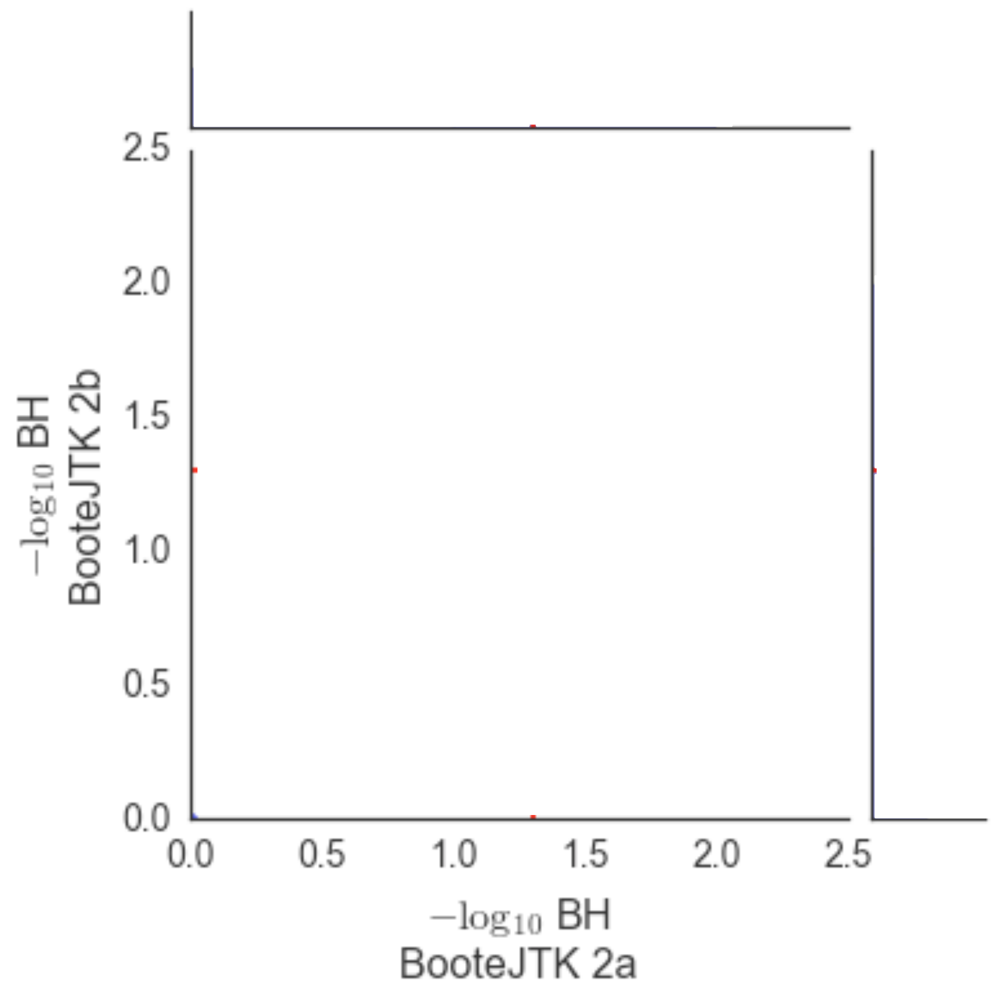




1: 0, 1, 2, 3, 4 ...
2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

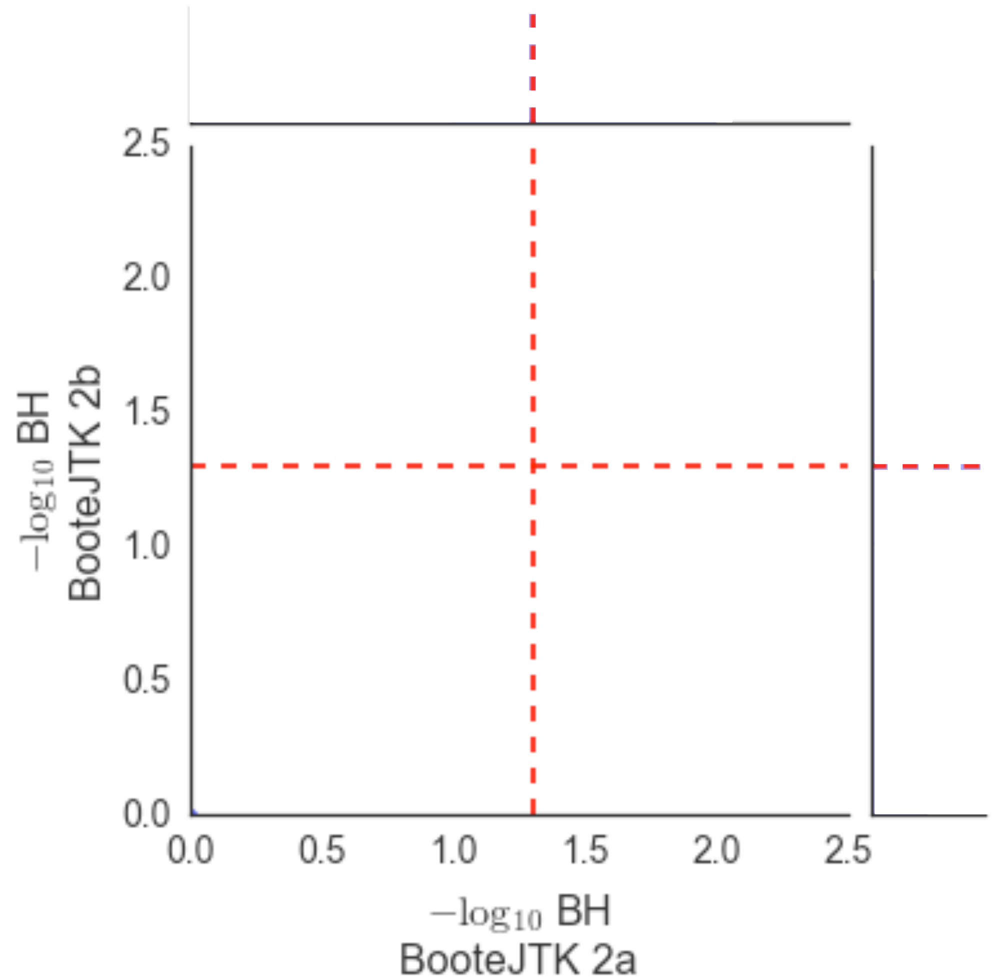
Hughes *et al.* (2009) “Harmonics of Circadian Gene Transcription in Mammals.”
PLoS Genetics, 2009. 5(4): e1000442

Comparison of downsampled dataset results



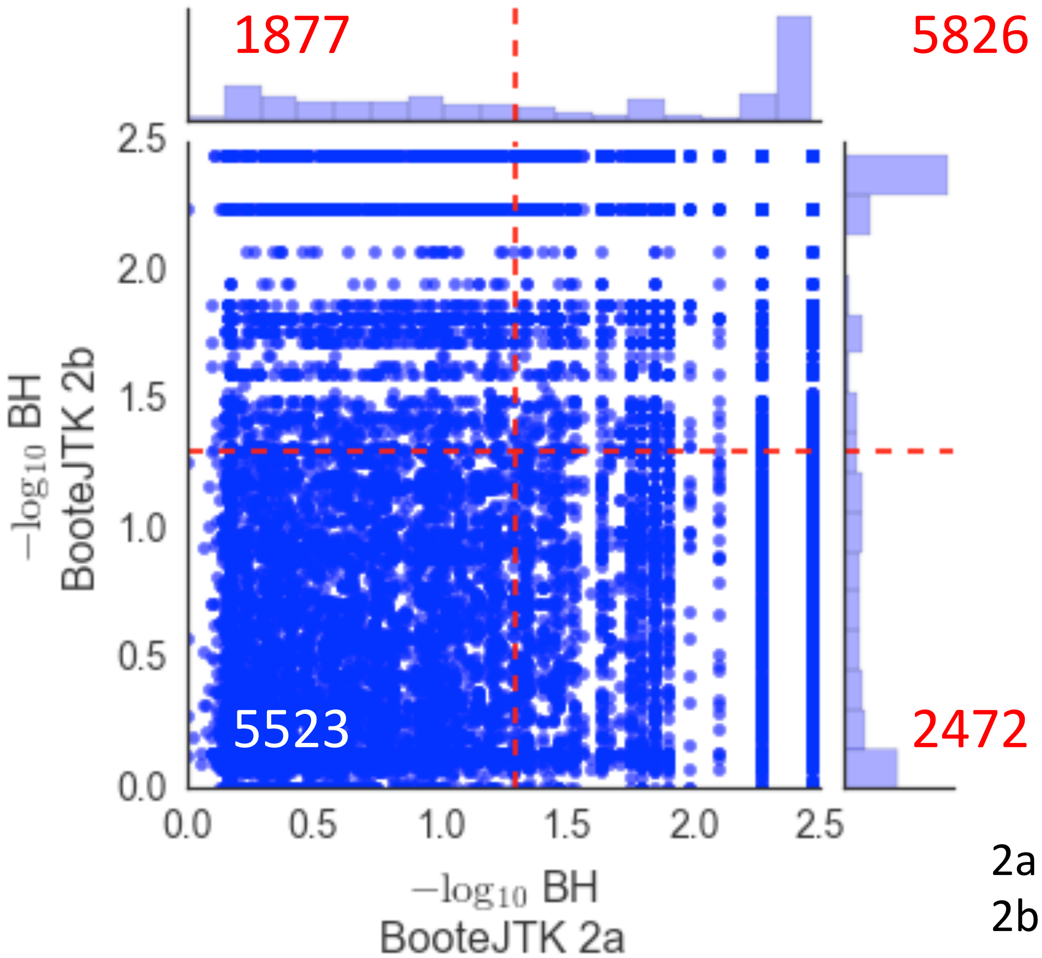
2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

Comparison of downsampled dataset results



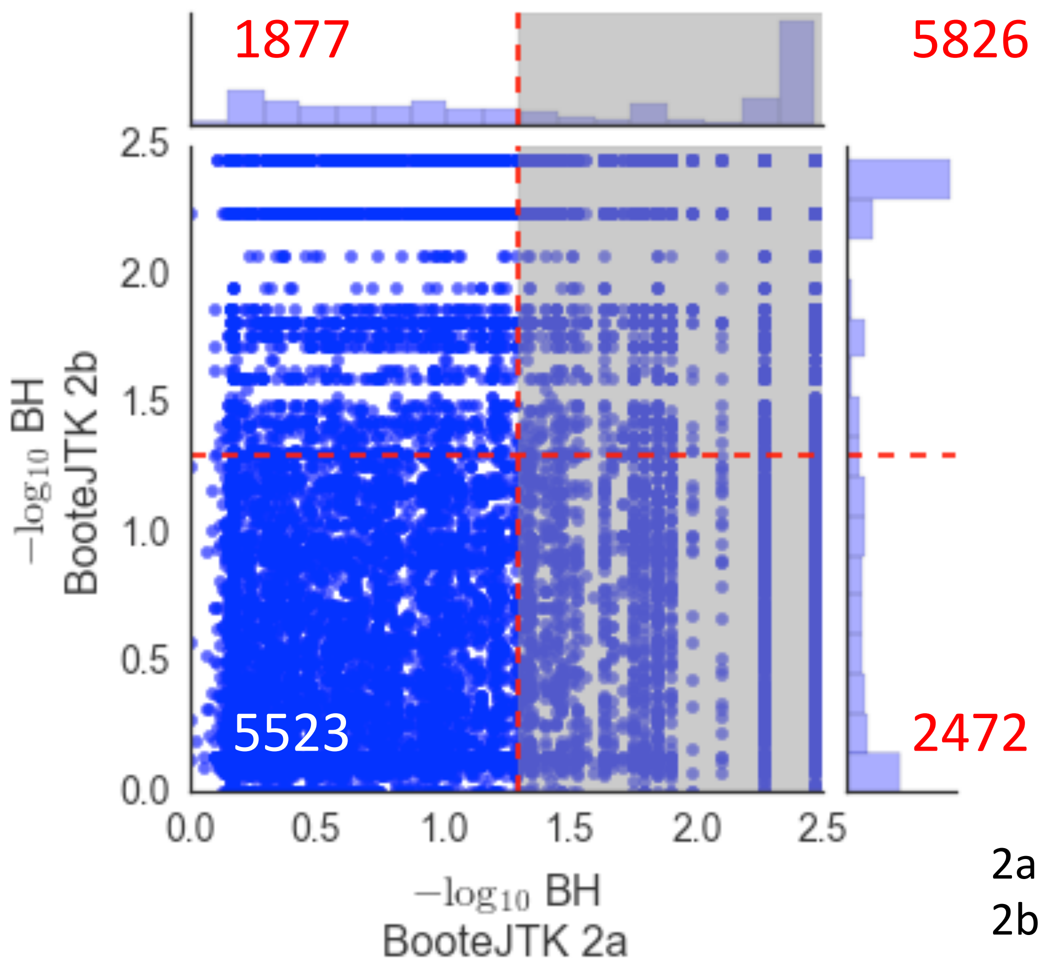
2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

Comparison of downsampled dataset results



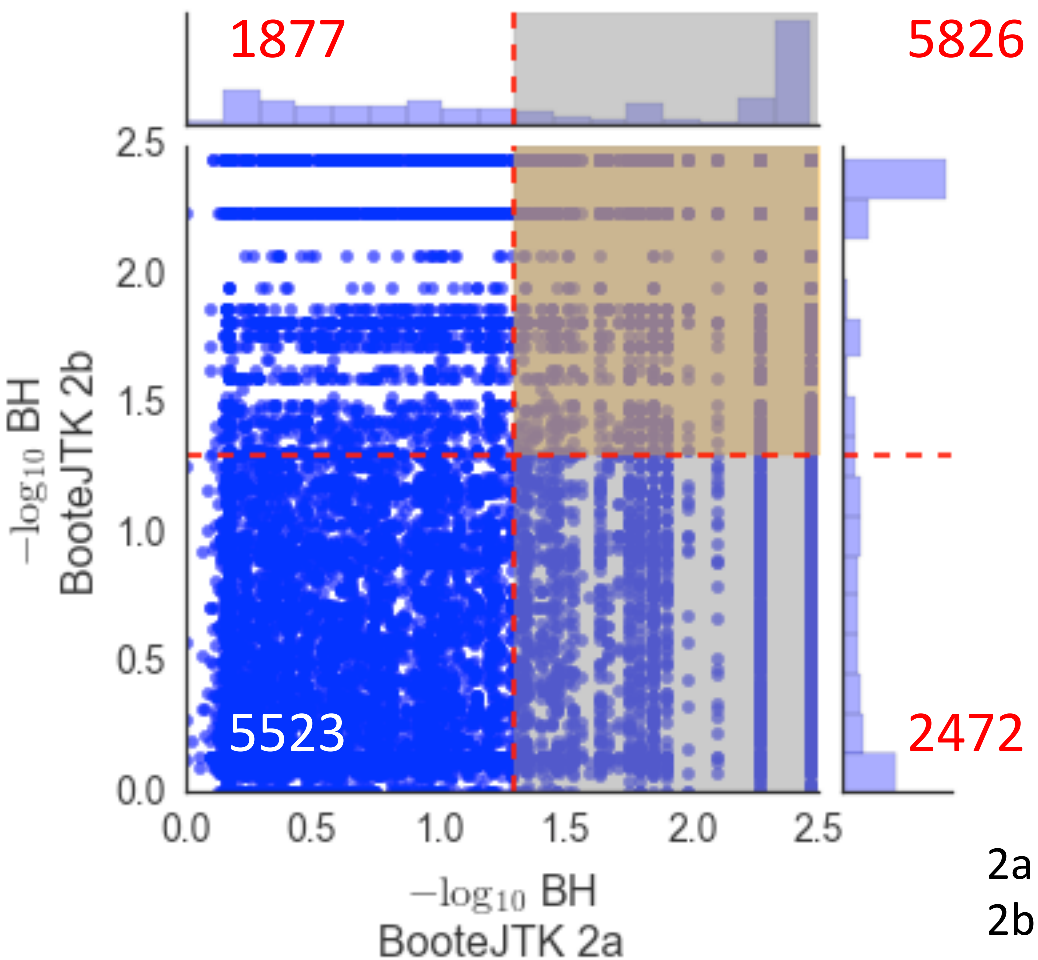
2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

Comparison of downsampled dataset results



2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

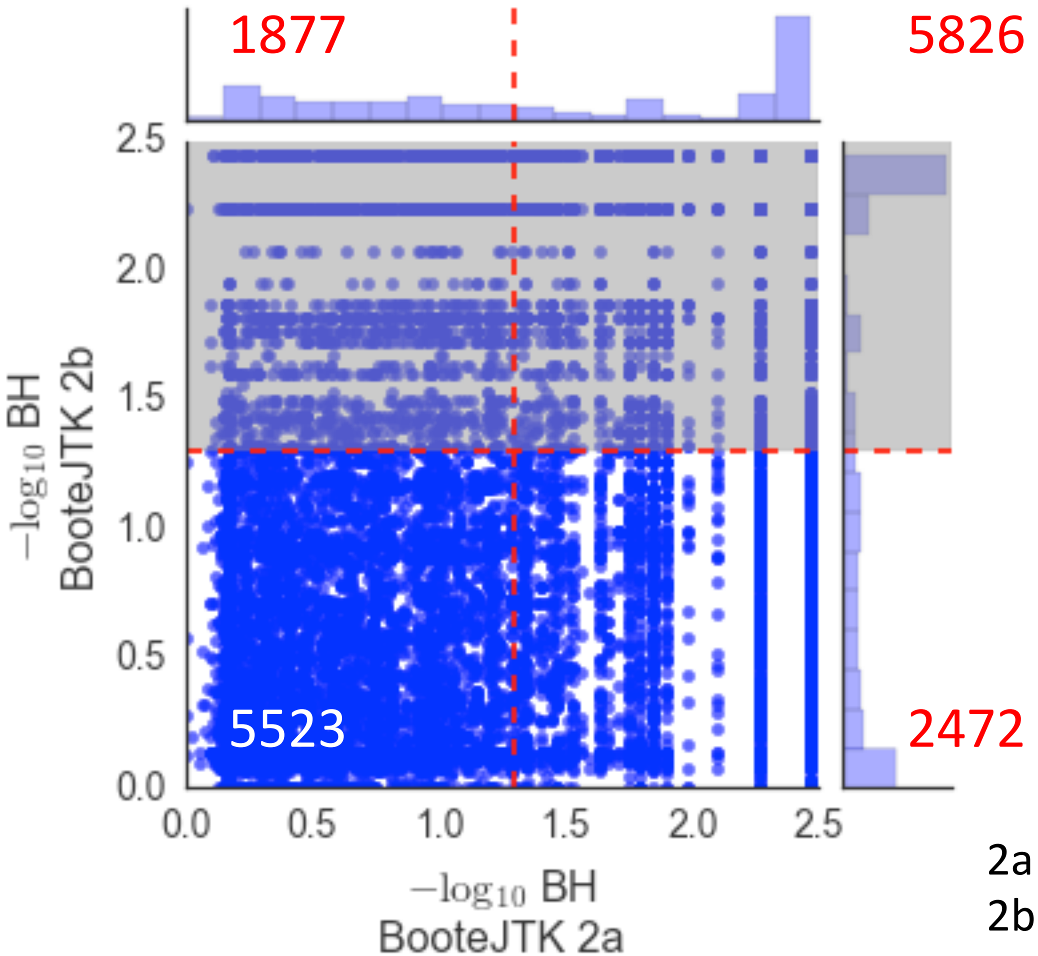
Comparison of downsampled dataset results



P(2b|2a) = 0.70

2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

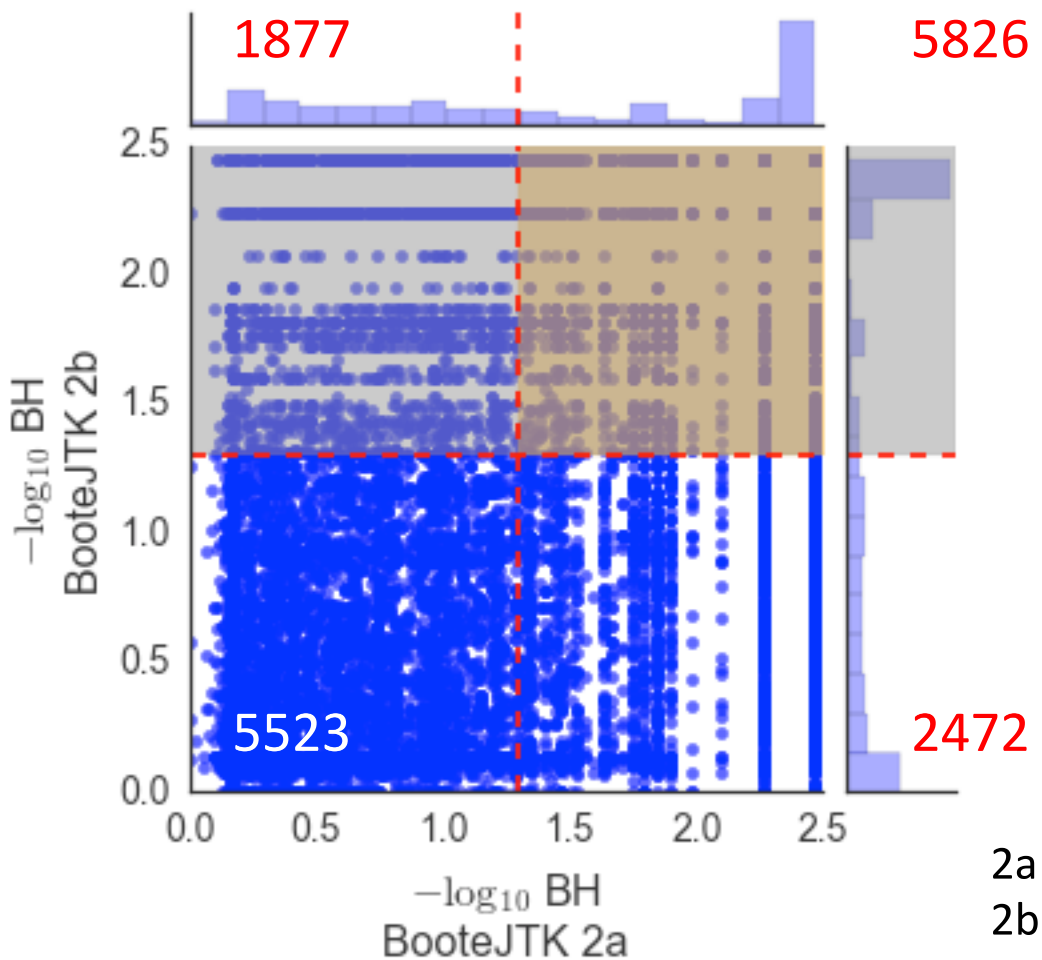
Comparison of downsampled dataset results



P(2b|2a) = 0.70

2a: 0, 2, 4, 6 ...
2b: 1, 3, 5, 7 ...

Comparison of downsampled dataset results



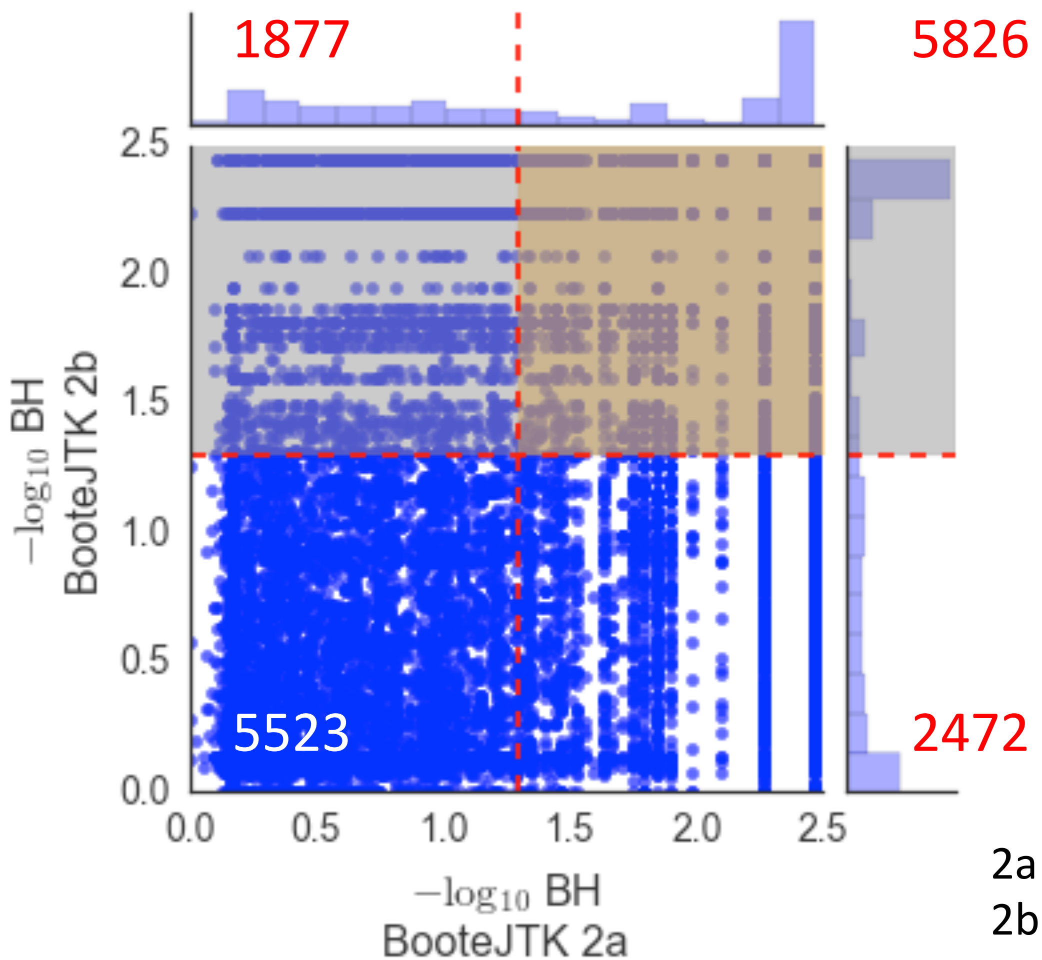
P(2b|2a) = 0.70

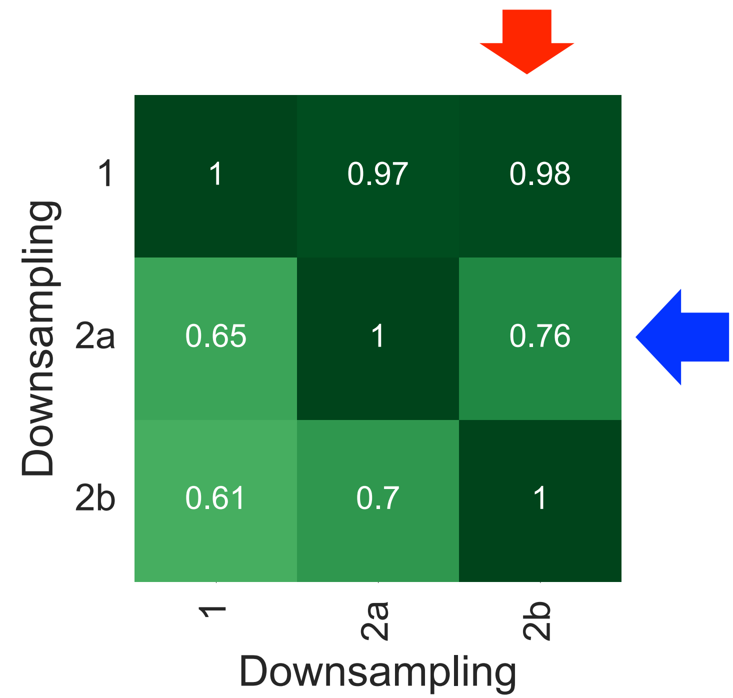
P(2a|2b) = 0.76

2a: 0, 2, 4, 6 ...

2b: 1, 3, 5, 7 ...

Comparison of downsampled dataset results





P(**row** | **column**)

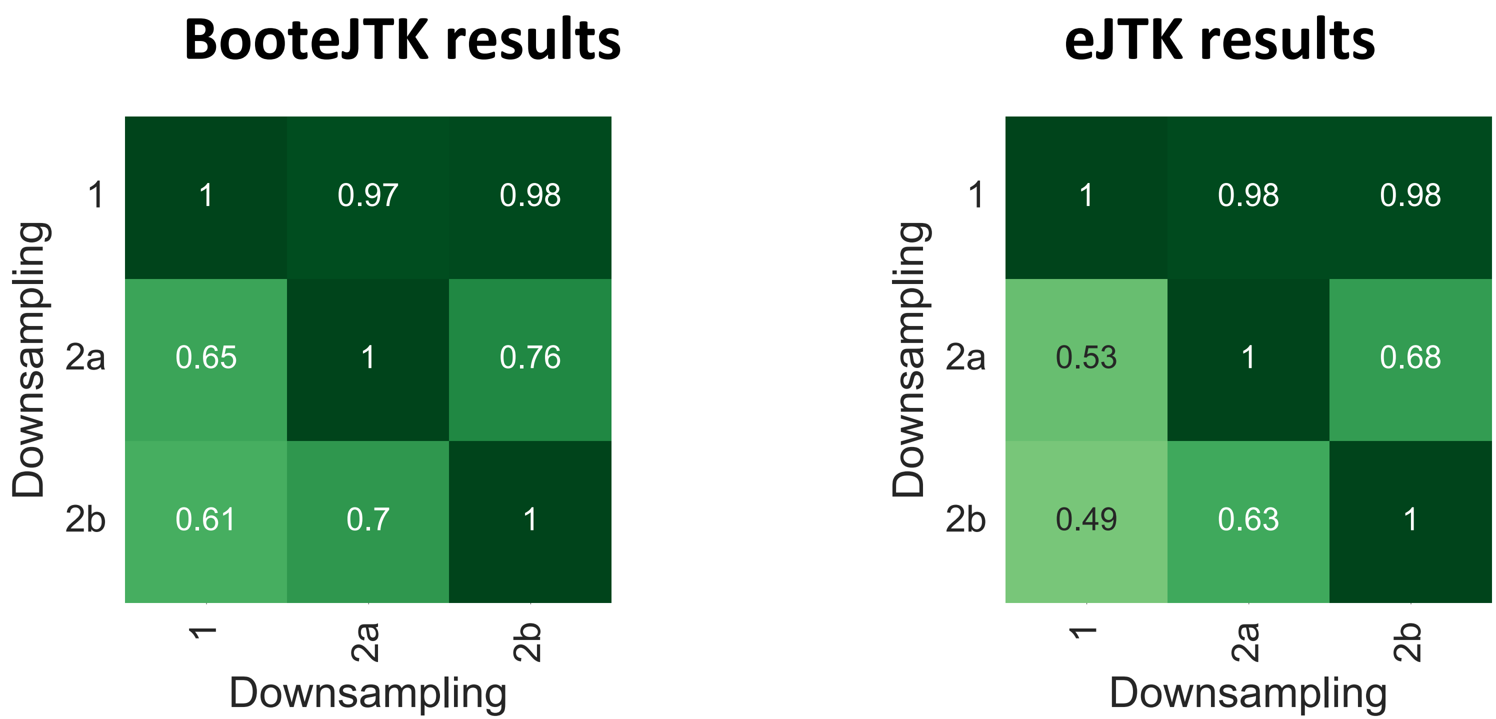
P(2b|2a) = 0.70

P(2a|2b) = 0.76

2a: 0, 2, 4, 6 ...

2b: 1, 3, 5, 7 ...

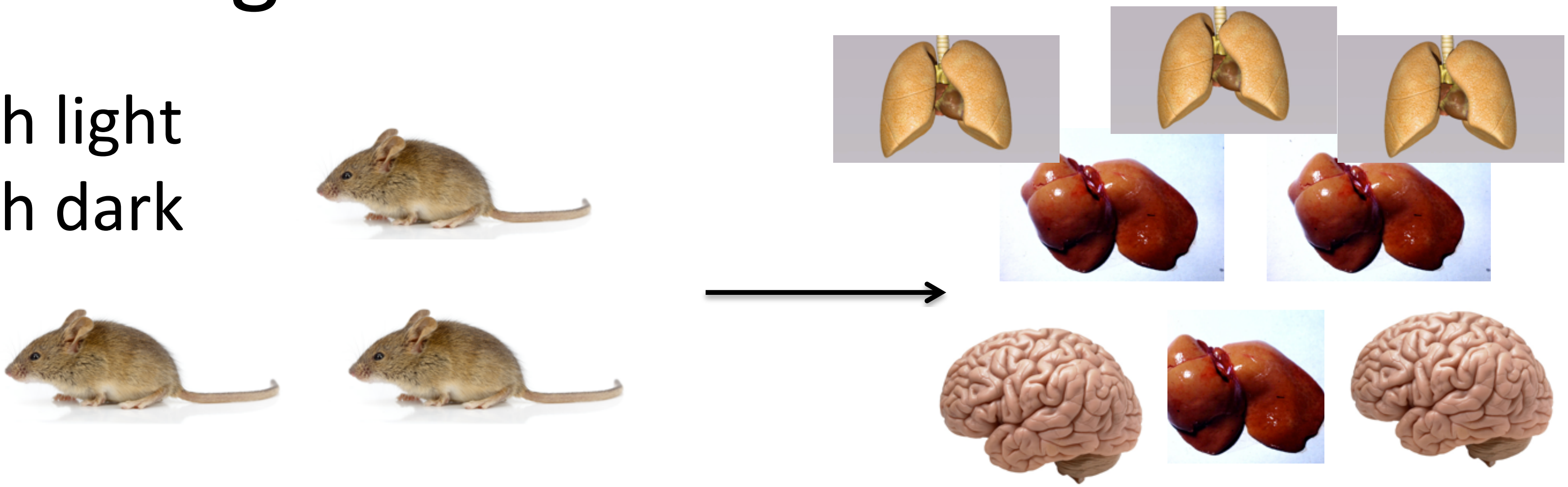
BooteJTK are more consistent as results are downsampled

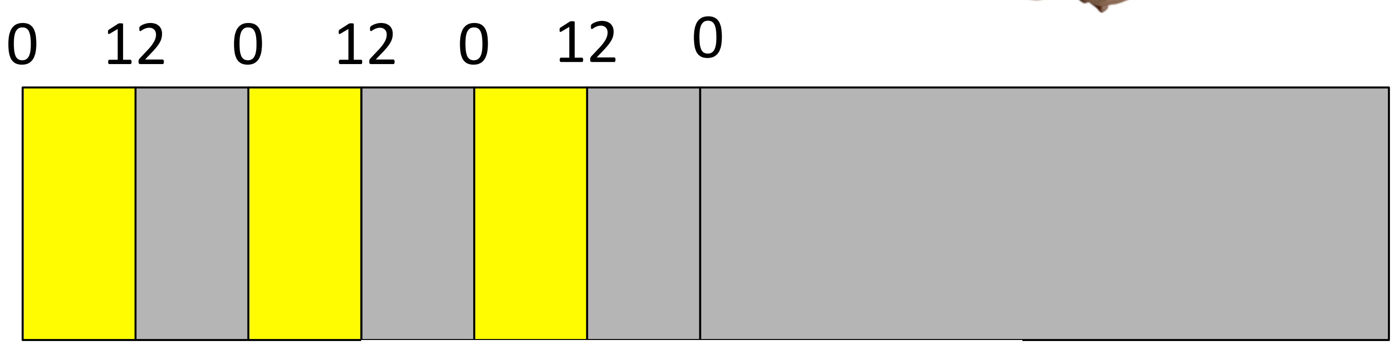


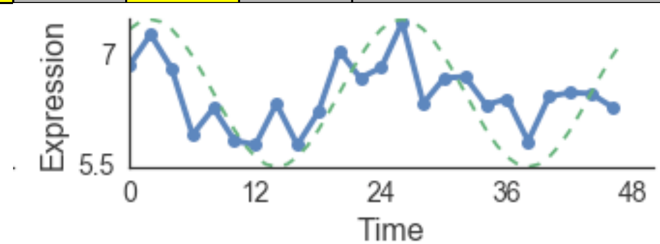
P(**row** | **column**)

Zhang *et al.* 2h 12 tissue dataset

12 h light
12 h dark

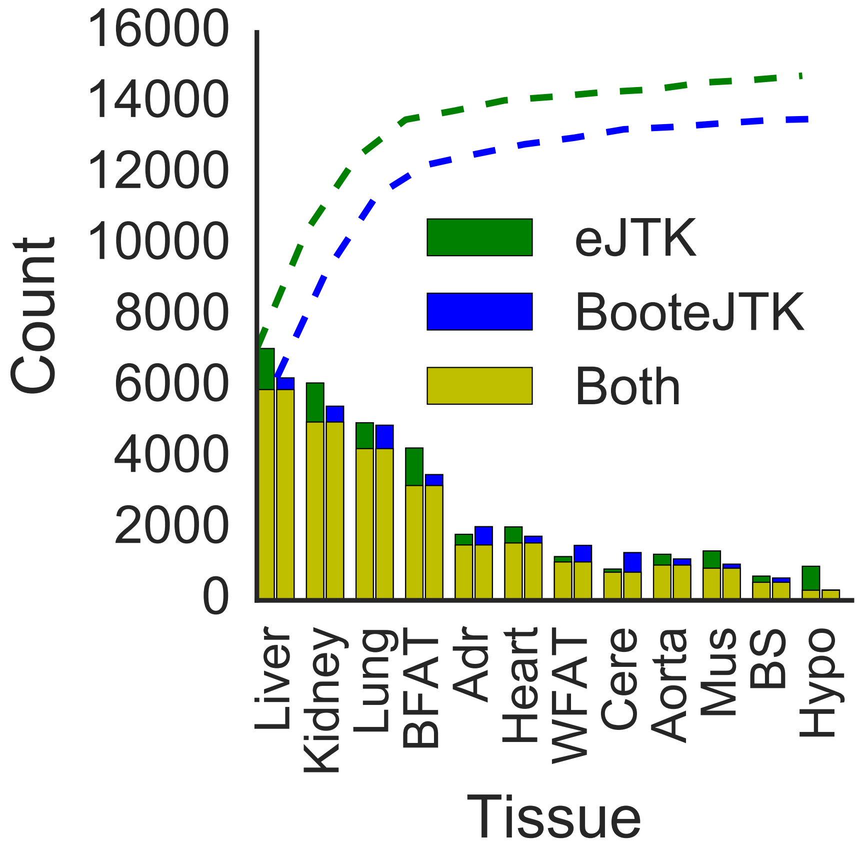




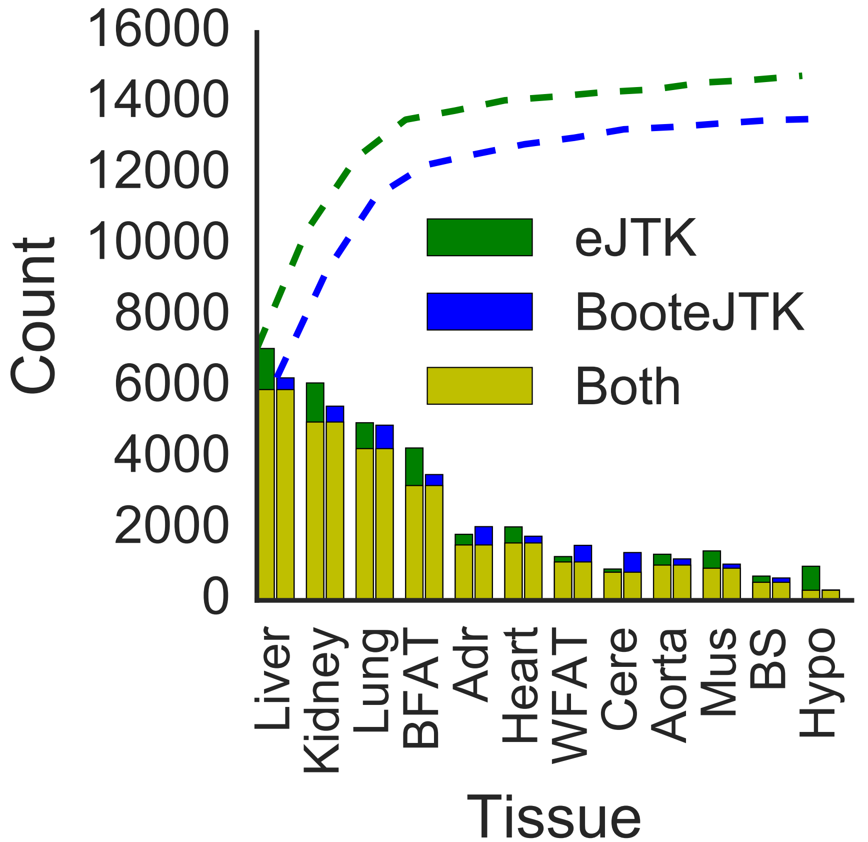


Zhang *et al.* (2014) "A circadian gene expression atlas in mammals: Implications for biology and medicine." PNAS (111) 45

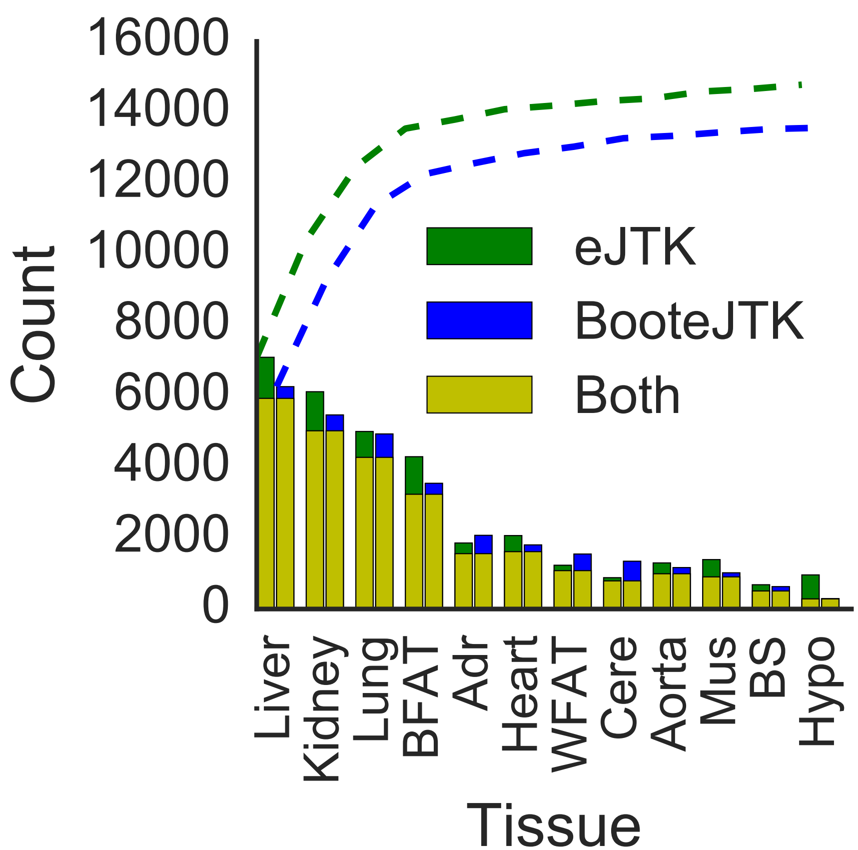
BooteJTK is more stringent than eJTK for most of the tissues

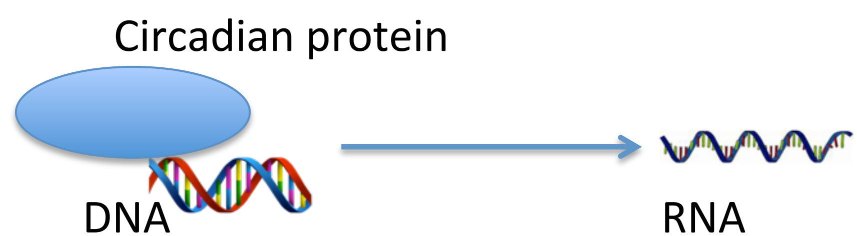


Is BooteJTK too stringent and missing rhythmic genes?



ChIP-Seq corroboration shows no decrease in core clock target enrichment in BooteJTK





Koike *et al*. (2012) "Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals" Vol. 338 p349-354

Overlap statistics show clusters of tissue types

BooteJTK results

	Adr	Lung	Kidney	Liver	WFAT	Aorta	BFAT	Heart	Mus	Hypo	BS	Cere
Adr	1	0.18	0.2	0.16	0.28	0.35	0.22	0.23	0.24	0.31	0.26	0.22
Lung	0.43	1	0.34	0.31	0.43	0.56	0.38	0.45	0.48	0.39	0.45	0.43
Kidney	0.52	0.38	1	0.4	0.48	0.51	0.41	0.43	0.43	0.5	0.45	0.41
Liver	0.5	0.39	0.45	1	0.5	0.49	0.45	0.45	0.43	0.5	0.39	0.42
WFAT	0.21	0.14	0.14	0.12	1	0.33	0.21	0.2	0.22	0.16	0.16	0.16
Aorta	0.19	0.13	0.11	0.09	0.25	1	0.22	0.18	0.21	0.15	0.16	0.15
BFAT	0.38	0.27	0.27	0.25	0.47	0.66	1	0.38	0.43	0.29	0.34	0.29
Heart	0.2	0.17	0.14	0.13	0.23	0.29	0.19	1	0.32	0.22	0.26	0.19
Mus	0.12	0.1	0.08	0.07	0.15	0.18	0.12	0.18	1	0.11	0.17	0.12
Hypo	0.04	0.02	0.03	0.02	0.03	0.04	0.02	0.03	0.03	1	0.12	0.07
BS	0.08	0.06	0.05	0.04	0.07	0.09	0.06	0.09	0.1	0.26	1	0.14
Cere	0.14	0.12	0.1	0.09	0.13	0.17	0.11	0.14	0.16	0.3	0.29	1

Brain tissue overlaps are a major difference between BooteJTK and eJTK results

BooteJTK results

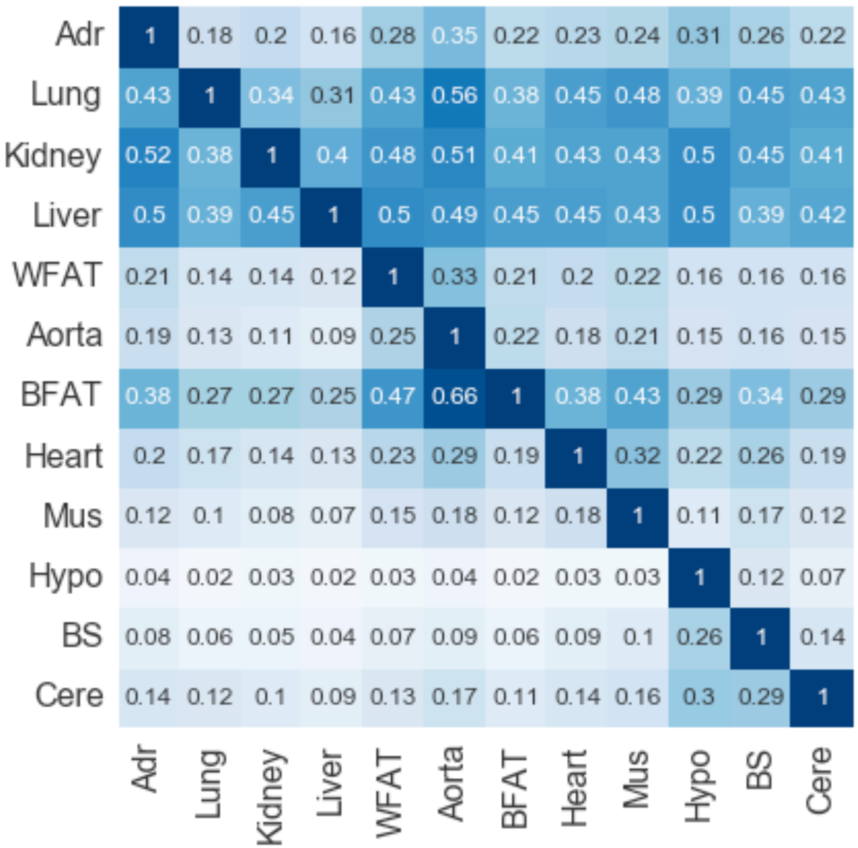
	Adr	Lung	Kidney	Liver	WFAT	Aorta	BFAT	Heart	Mus	Hypo	BS	Cere
Adr	1	0.18	0.2	0.16	0.28	0.35	0.22	0.23	0.24	0.31	0.26	0.22
Lung	0.43	1	0.34	0.31	0.43	0.56	0.38	0.45	0.48	0.39	0.45	0.43
Kidney	0.52	0.38	1	0.4	0.48	0.51	0.41	0.43	0.43	0.5	0.45	0.41
Liver	0.5	0.39	0.45	1	0.5	0.49	0.45	0.45	0.43	0.5	0.39	0.42
WFAT	0.21	0.14	0.14	0.12	1	0.33	0.21	0.2	0.22	0.16	0.16	0.16
Aorta	0.19	0.13	0.11	0.09	0.25	1	0.22	0.18	0.21	0.15	0.16	0.15
BFAT	0.38	0.27	0.27	0.25	0.47	0.66	1	0.38	0.43	0.29	0.34	0.29
Heart	0.2	0.17	0.14	0.13	0.23	0.29	0.19	1	0.32	0.22	0.26	0.19
Mus	0.12	0.1	0.08	0.07	0.15	0.18	0.12	0.18	1	0.11	0.17	0.12
Hypo	0.04	0.02	0.03	0.02	0.03	0.04	0.02	0.03	0.03	1	0.12	0.07
BS	0.08	0.06	0.05	0.04	0.07	0.09	0.06	0.09	0.1	0.26	1	0.14
Cere	0.14	0.12	0.1	0.09	0.13	0.17	0.11	0.14	0.16	0.3	0.29	1

eJTK results

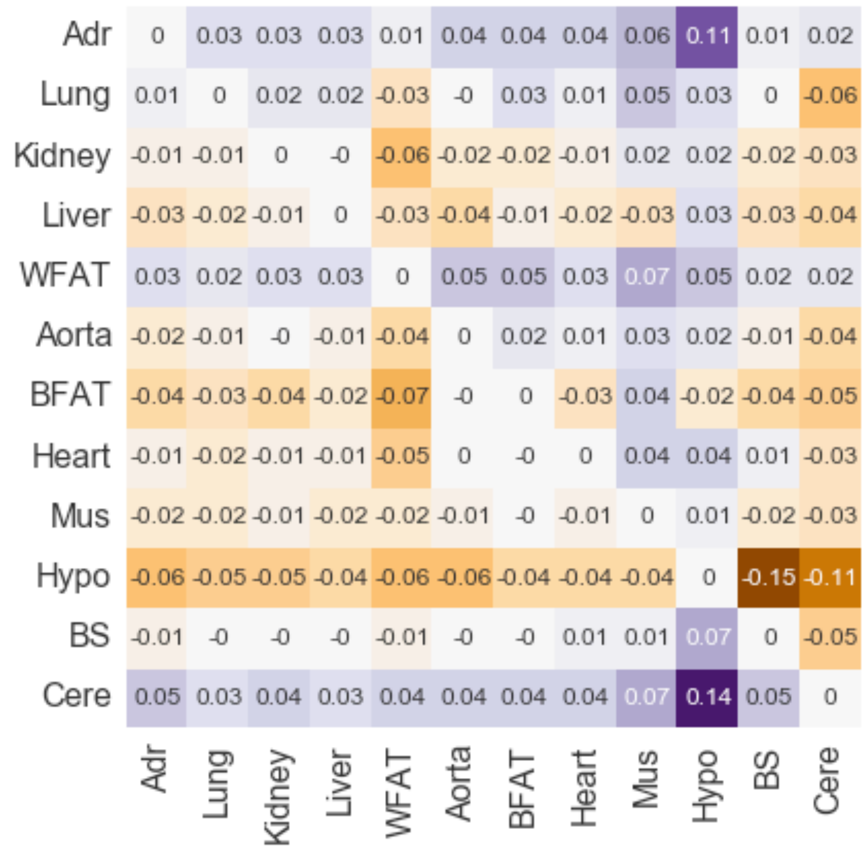
	Adr	Lung	Kidney	Liver	WFAT	Aorta	BFAT	Heart	Mus	Hypo	BS	Cere
Adr	1	0.16	0.16	0.14	0.27	0.3	0.18	0.19	0.18	0.2	0.25	0.2
Lung	0.42	1	0.32	0.29	0.46	0.56	0.35	0.44	0.43	0.37	0.45	0.49
Kidney	0.53	0.39	1	0.4	0.55	0.53	0.43	0.44	0.41	0.48	0.48	0.45
Liver	0.52	0.41	0.46	1	0.53	0.53	0.45	0.46	0.45	0.47	0.42	0.46
WFAT	0.18	0.11	0.11	0.09	1	0.27	0.15	0.17	0.15	0.11	0.14	0.14
Aorta	0.21	0.15	0.11	0.1	0.29	1	0.2	0.18	0.18	0.13	0.17	0.19
BFAT	0.42	0.3	0.3	0.27	0.54	0.66	1	0.41	0.39	0.31	0.39	0.33
Heart	0.21	0.18	0.15	0.14	0.28	0.28	0.2	1	0.28	0.17	0.25	0.22
Mus	0.13	0.12	0.09	0.09	0.17	0.19	0.13	0.19	1	0.1	0.19	0.15
Hypo	0.11	0.07	0.08	0.06	0.09	0.1	0.07	0.08	0.07	1	0.27	0.18
BS	0.09	0.06	0.05	0.04	0.08	0.09	0.06	0.08	0.09	0.19	1	0.19
Cere	0.09	0.09	0.06	0.06	0.1	0.13	0.07	0.09	0.09	0.16	0.24	1

Brain tissue overlaps are a major difference between BooteJTK and eJTK results

BooteJTK results



BooteJTK - eJTK results



eJTK overlap > BooteJTK overlap

eJTK overlap < BooteJTK overlap

Brain tissue overlaps are a major difference between BooteJTK and eJTK results

BooteJTK results

	Adr	Lung	Kidney	Liver	WFAT	Aorta	BFAT	Heart	Mus	Hypo	BS	Cere
Adr	1	0.18	0.2	0.16	0.28	0.35	0.22	0.23	0.24	0.31	0.26	0.22
Lung	0.43	1	0.34	0.31	0.43	0.56	0.38	0.45	0.48	0.39	0.45	0.43
Kidney	0.52	0.38	1	0.4	0.48	0.51	0.41	0.43	0.43	0.5	0.45	0.41
Liver	0.5	0.39	0.45	1	0.5	0.49	0.45	0.45	0.43	0.5	0.39	0.42
WFAT	0.21	0.14	0.14	0.12	1	0.33	0.21	0.2	0.22	0.16	0.16	0.16
Aorta	0.19	0.13	0.11	0.09	0.25	1	0.22	0.18	0.21	0.15	0.16	0.15
BFAT	0.38	0.27	0.27	0.25	0.47	0.66	1	0.38	0.43	0.29	0.34	0.29
Heart	0.2	0.17	0.14	0.13	0.23	0.29	0.19	1	0.32	0.22	0.26	0.19
Mus	0.12	0.1	0.08	0.07	0.15	0.18	0.12	0.18	1	0.11	0.17	0.12
Hypo	0.04	0.02	0.03	0.02	0.03	0.04	0.02	0.03	0.03	1	0.12	0.07
BS	0.08	0.06	0.05	0.04	0.07	0.09	0.06	0.09	0.1	0.26	1	0.14
Cere	0.14	0.12	0.1	0.09	0.13	0.17	0.11	0.14	0.16	0.3	0.29	1

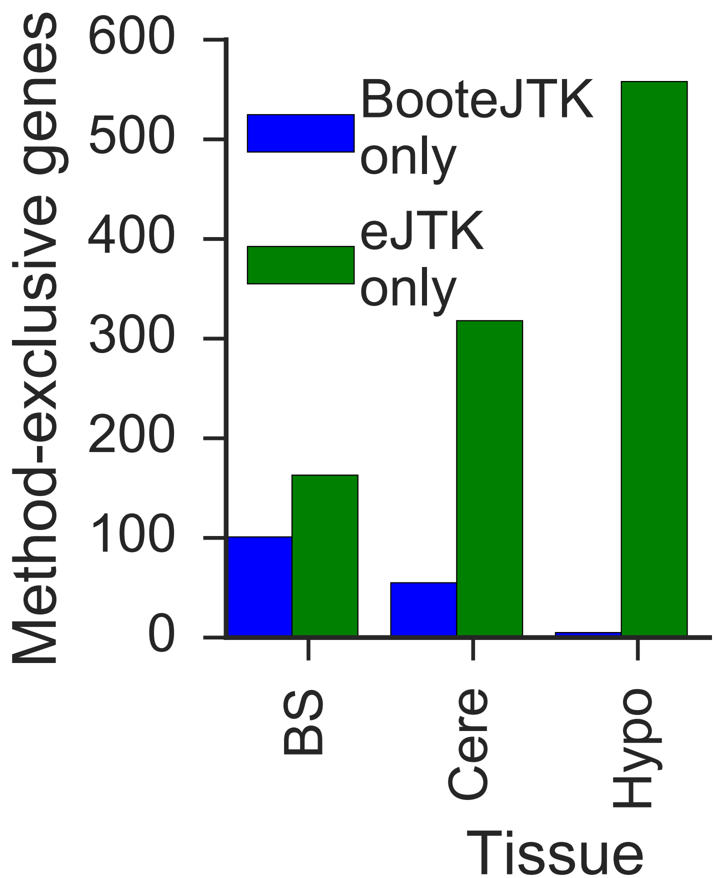
BooteJTK - eJTK results

	Adr	Lung	Kidney	Liver	WFAT	Aorta	BFAT	Heart	Mus	Hypo	BS	Cere
Adr	0	0.03	0.03	0.03	0.01	0.04	0.04	0.04	0.06	0.11	0.01	0.02
Lung	0.01	0	0.02	0.02	-0.03	-0	0.03	0.01	0.05	0.03	0	-0.06
Kidney	-0.01	-0.01	0	-0	-0.06	-0.02	-0.02	-0.01	0.02	0.02	-0.02	-0.03
Liver	-0.03	-0.02	-0.01	0	-0.03	-0.04	-0.01	-0.02	-0.03	0.03	-0.03	-0.04
WFAT	0.03	0.02	0.03	0.03	0	0.05	0.05	0.03	0.07	0.05	0.02	0.02
Aorta	-0.02	-0.01	-0	-0.01	-0.04	0	0.02	0.01	0.03	0.02	-0.01	-0.04
BFAT	-0.04	-0.03	-0.04	-0.02	-0.07	-0	0	-0.03	0.04	-0.02	-0.04	-0.05
Heart	-0.01	-0.02	-0.01	-0.01	-0.05	0	-0	0	0.04	0.04	0.01	-0.03
Mus	-0.02	-0.02	-0.01	-0.02	-0.02	-0.01	-0	-0.01	0	0.01	-0.02	-0.03
Hypo	-0.06	-0.05	-0.05	-0.04	-0.06	-0.06	-0.04	-0.04	-0.04	0	-0.15	-0.11
BS	-0.01	-0	-0	-0	-0.01	-0	-0	0.01	0.01	0.07	0	-0.05
Cere	0.05	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.07	0.14	0.05	0

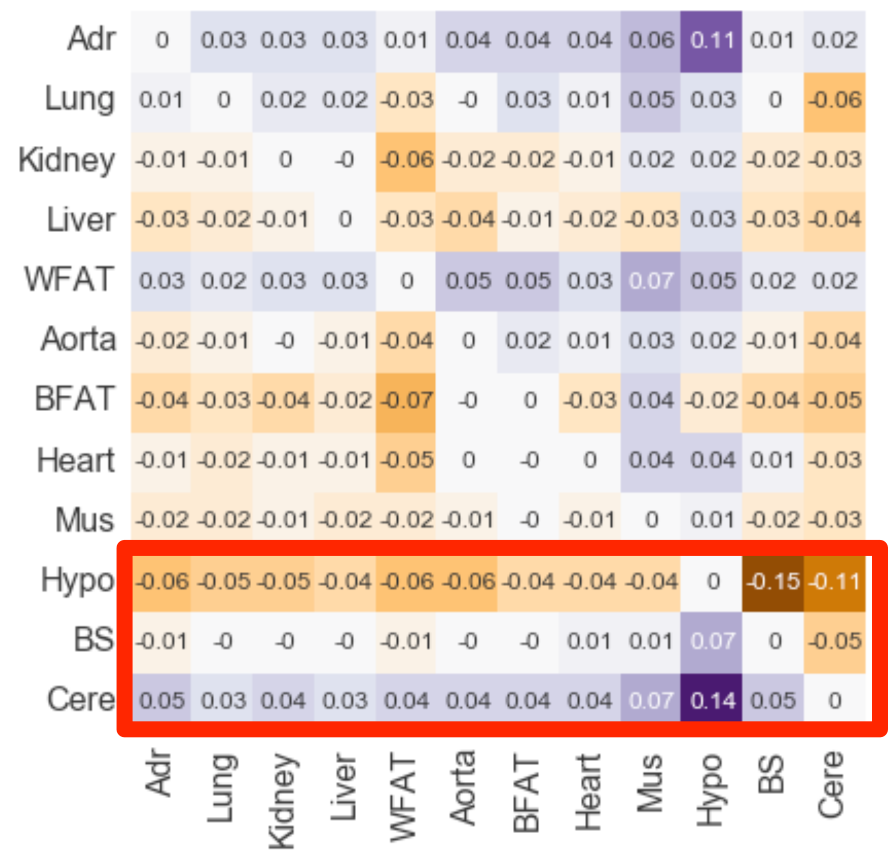
eJTK overlap > BooteJTK overlap

eJTK overlap < BooteJTK overlap

Brain tissue overlaps are a major difference between BooteJTK and eJTK results



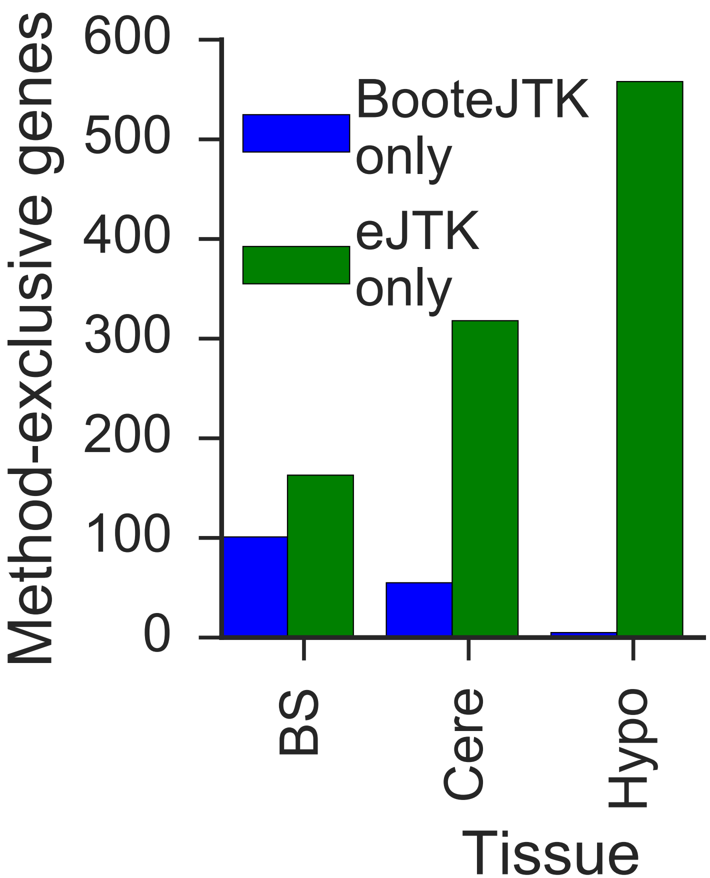
BooteJTK - eJTK results

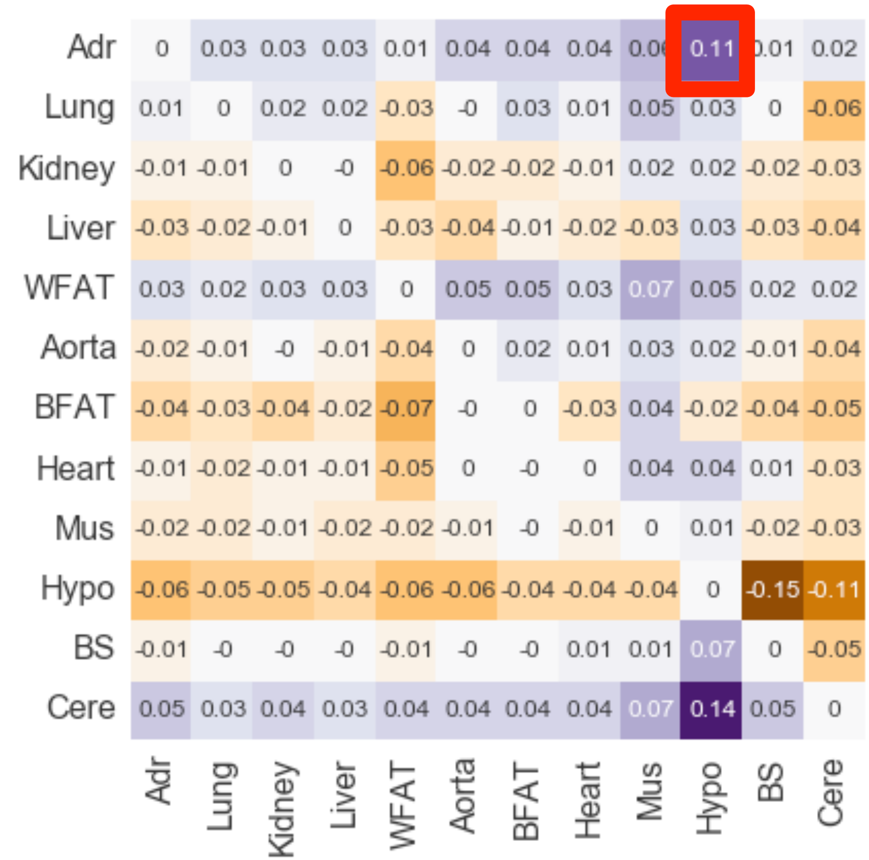


eJTK overlap > BooteJTK overlap

eJTK overlap < BooteJTK overlap

Adrenal-Hypothalamus rhythmic overlap is a large difference between BooteJTK and eJTK results

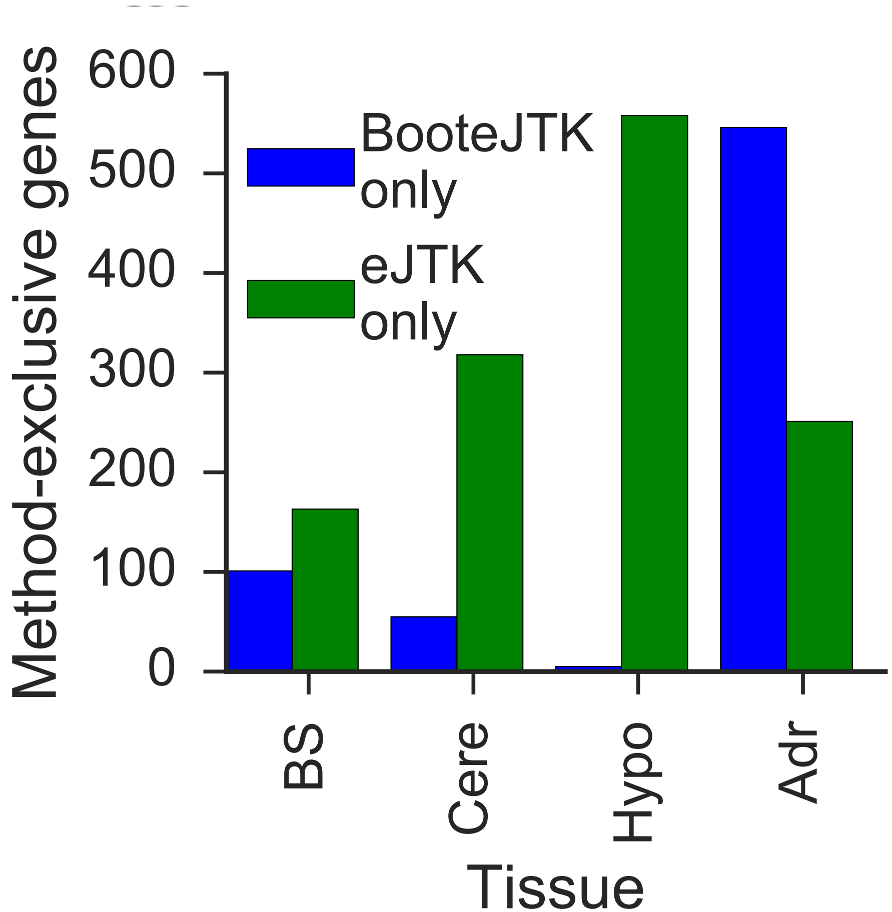


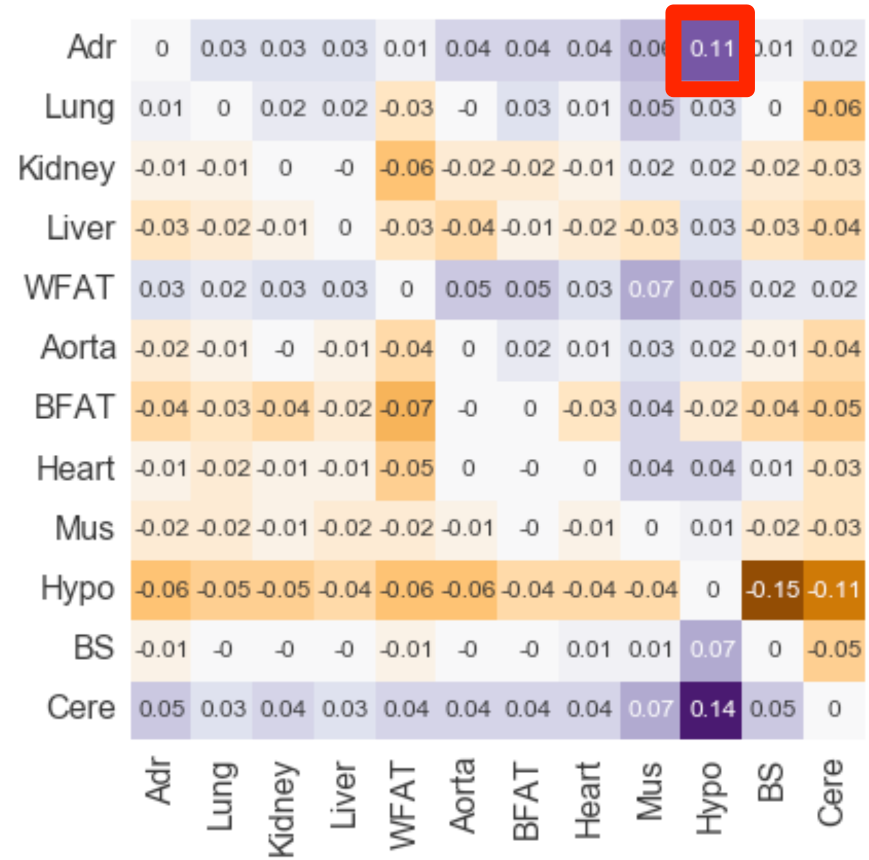


eJTK overlap > BooteJTK overlap

eJTK overlap < BooteJTK overlap

Adrenal-Hypothalamus rhythmic overlap is a large difference between BooteJTK and eJTK results

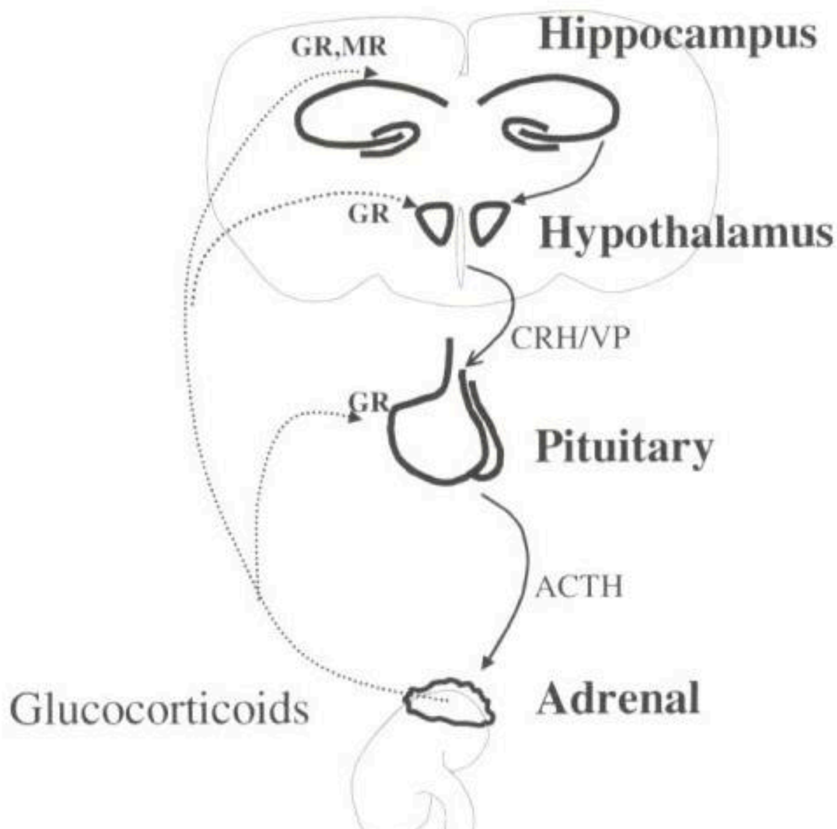




eJTK overlap > BooteJTK overlap

eJTK overlap < BooteJTK overlap

The hypothalamus and adrenals are involved in an endrocrine feedback loop



en.wikipedia.org/wiki/Hypothalamic–pituitary–adrenal_axis

	Adr	Lung	Kidney	Liver	WFAT	Aorta	BFAT	Heart	Mus	Hypo	BS	Cere
Adr	0	0.03	0.03	0.03	0.01	0.04	0.04	0.04	0.06	0.11	0.01	0.02
Lung	0.01	0	0.02	0.02	-0.03	-0	0.03	0.01	0.05	0.03	0	-0.06
Kidney	-0.01	-0.01	0	-0	-0.06	-0.02	-0.02	-0.01	0.02	0.02	-0.02	-0.03
Liver	-0.03	-0.02	-0.01	0	-0.03	-0.04	-0.01	-0.02	-0.03	0.03	-0.03	-0.04
WFAT	0.03	0.02	0.03	0.03	0	0.05	0.05	0.03	0.07	0.05	0.02	0.02
Aorta	-0.02	-0.01	-0	-0.01	-0.04	0	0.02	0.01	0.03	0.02	-0.01	-0.04
BFAT	-0.04	-0.03	-0.04	-0.02	-0.07	-0	0	-0.03	0.04	-0.02	-0.04	-0.05
Heart	-0.01	-0.02	-0.01	-0.01	-0.05	0	-0	0	0.04	0.04	0.01	-0.03
Mus	-0.02	-0.02	-0.01	-0.02	-0.02	-0.01	-0	-0.01	0	0.01	-0.02	-0.03
Hypo	-0.06	-0.05	-0.05	-0.04	-0.06	-0.06	-0.04	-0.04	-0.04	0	-0.15	-0.11
BS	-0.01	-0	-0	-0	-0.01	-0	-0	0.01	0.01	0.07	0	-0.05
Cere	0.05	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.07	0.14	0.05	0

eJTK overlap > BooteJTK overlap

eJTK overlap < BooteJTK overlap

Outline

- Biological and Statistical Background
- Improvements to JTK_CYCLE
  - Empirical JTK_CYCLE (eJTK)
  - **Bootstrap eJTK (BooteJTK)**
    - **Greater consistency than eJTK**
    - **More stringent than eJTK**
    - **Differences in results are biologically supported**

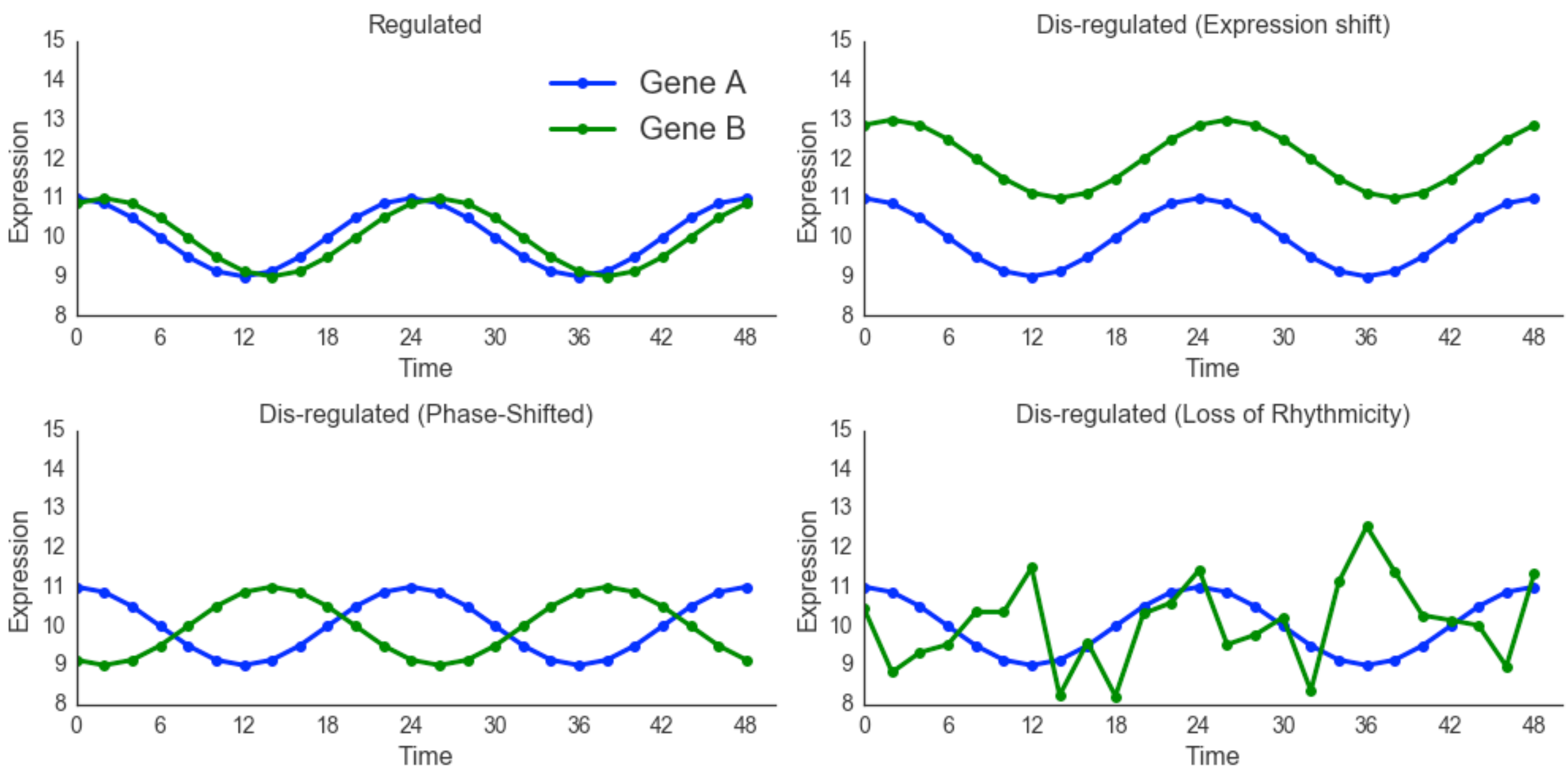
Outline

- Biological and Statistical Background
- Improvements to JTK_CYCLE
- **Comparing rhythmicity across conditions**

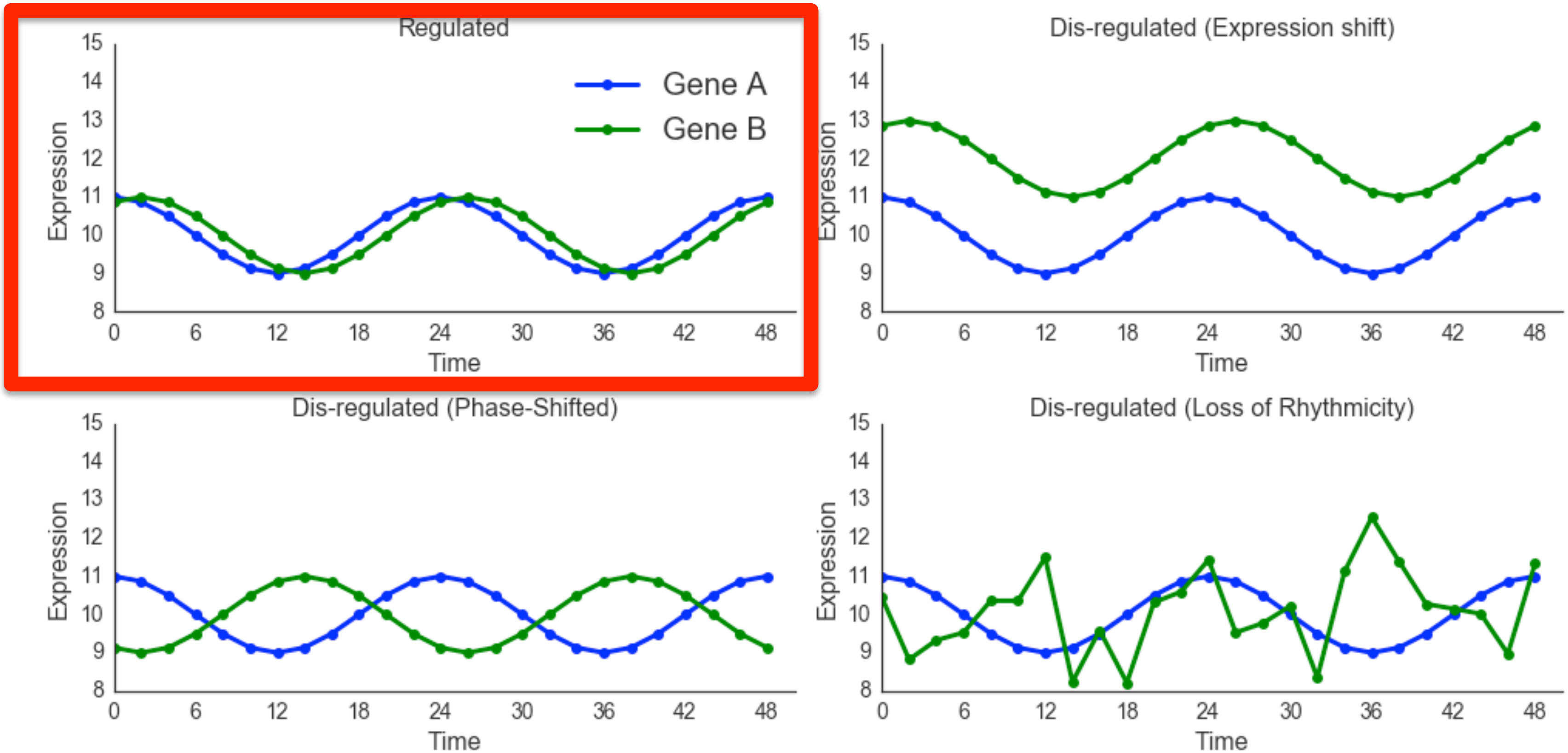
Outline

- Biological and Statistical Background
- Improvements to JTK_CYCLE
- **Comparing rhythmicity across conditions**
  - **A method that produces accurate p-values for differential rhythmicity**

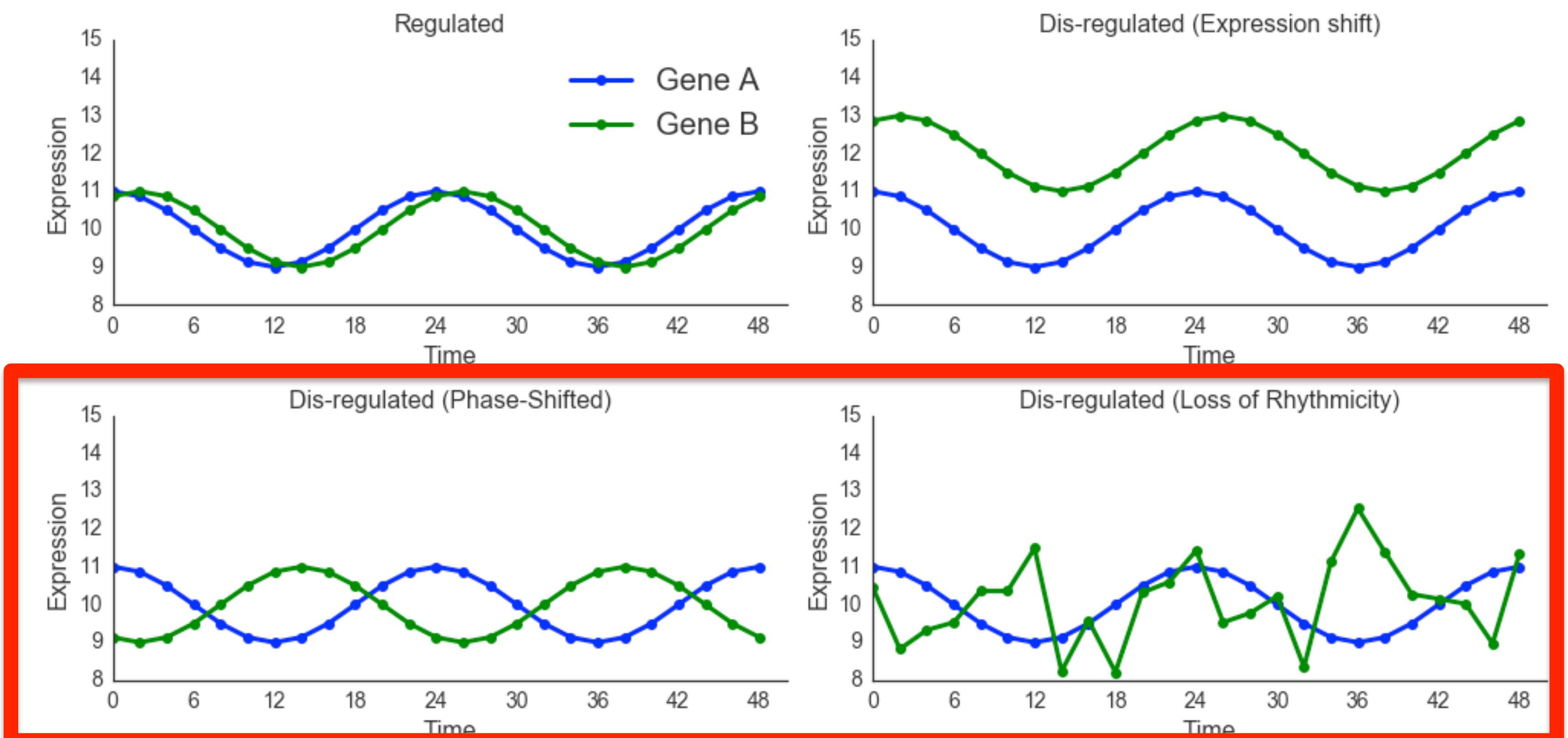
Dis-regulation can be more than changes in expression level



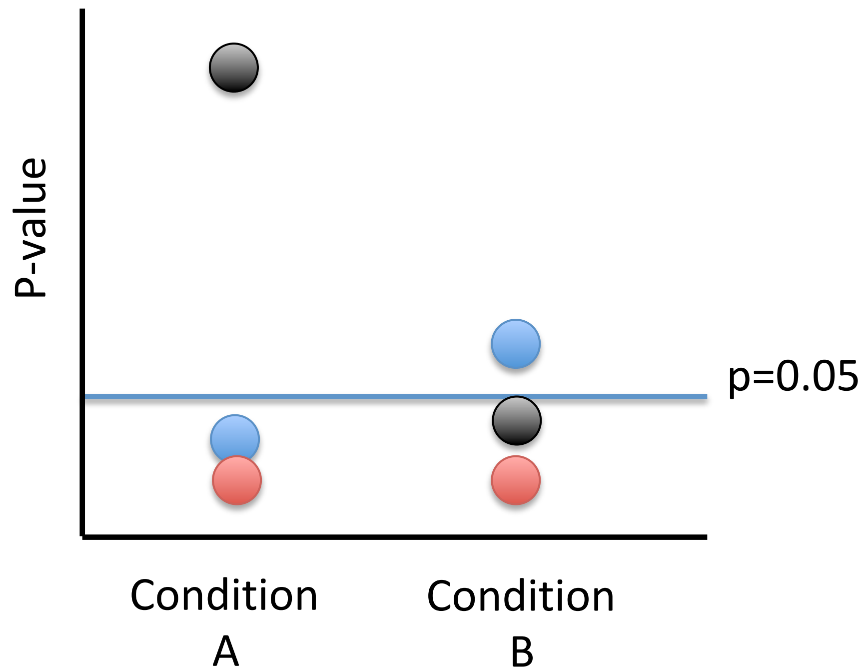
Dis-regulation can be more than changes in expression level



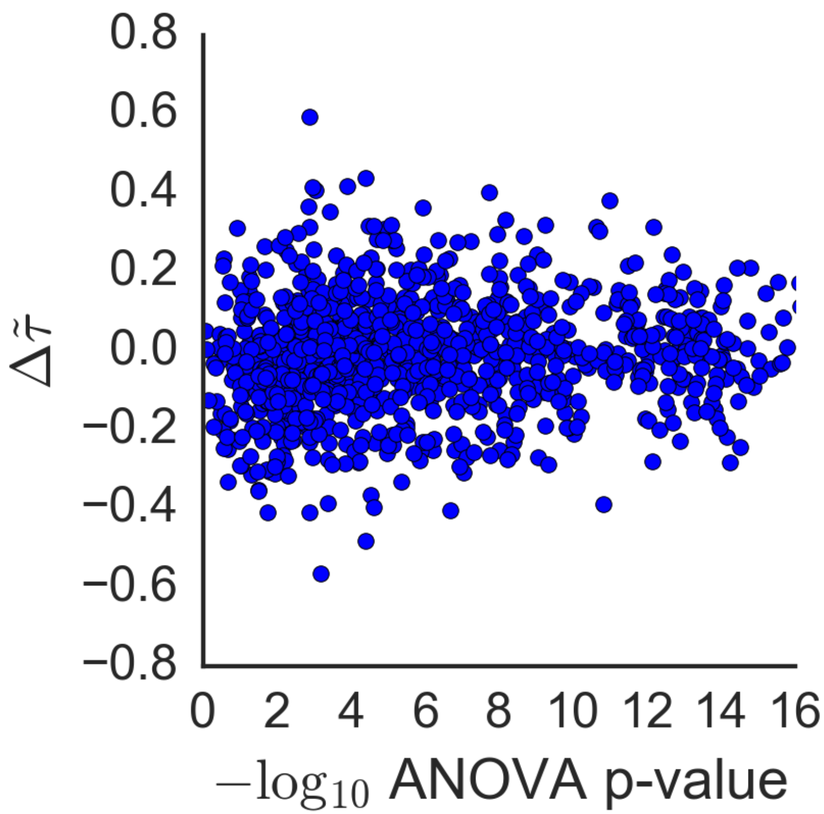
Dis-regulation can be more than changes in expression level

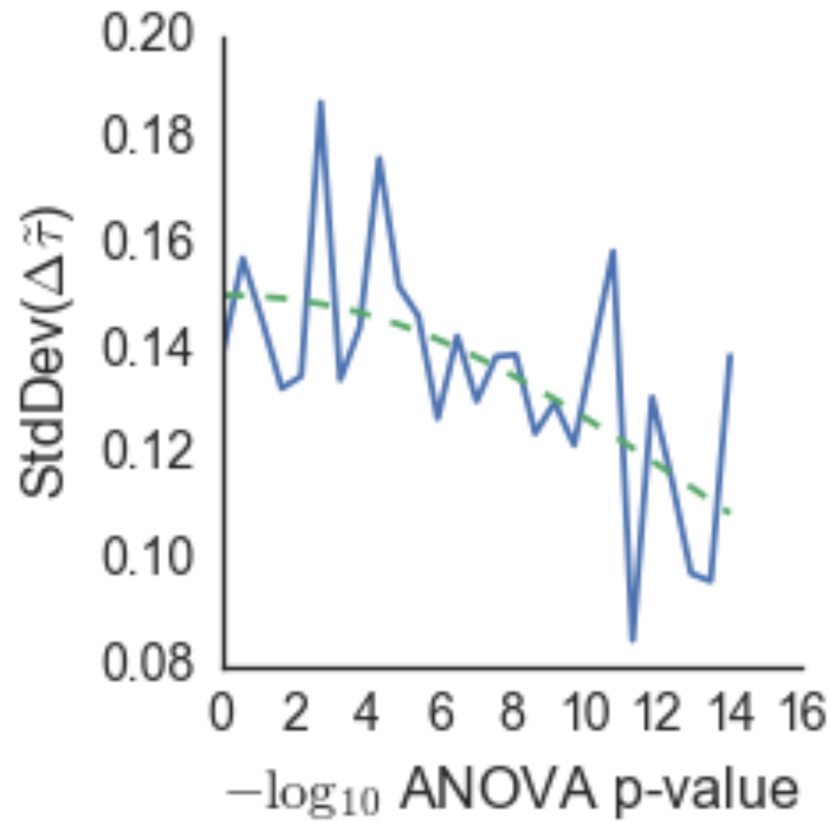


Previous approach only looks at p-values relative to threshold

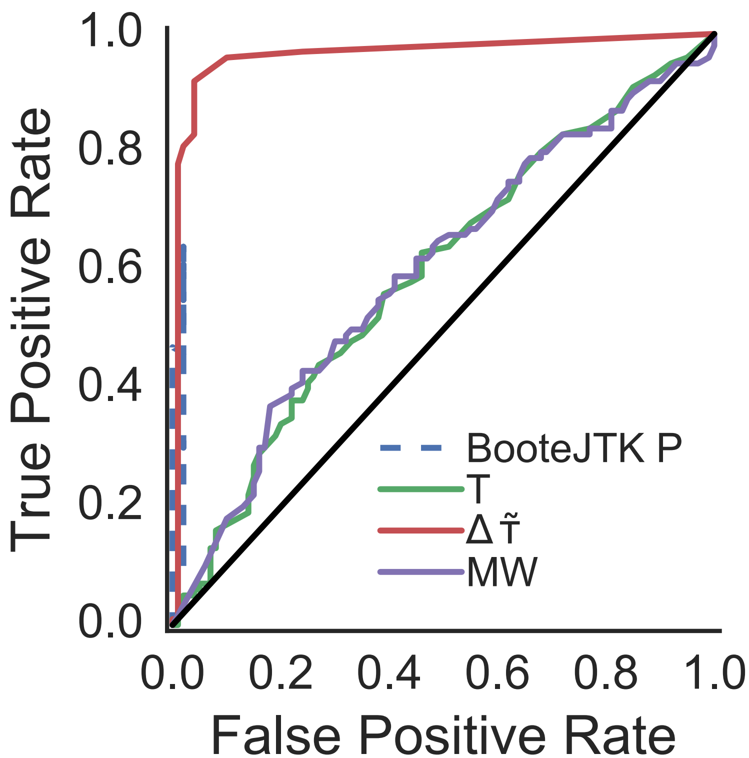


We can estimate the variance in our Tau score based on the noisiness of the time series

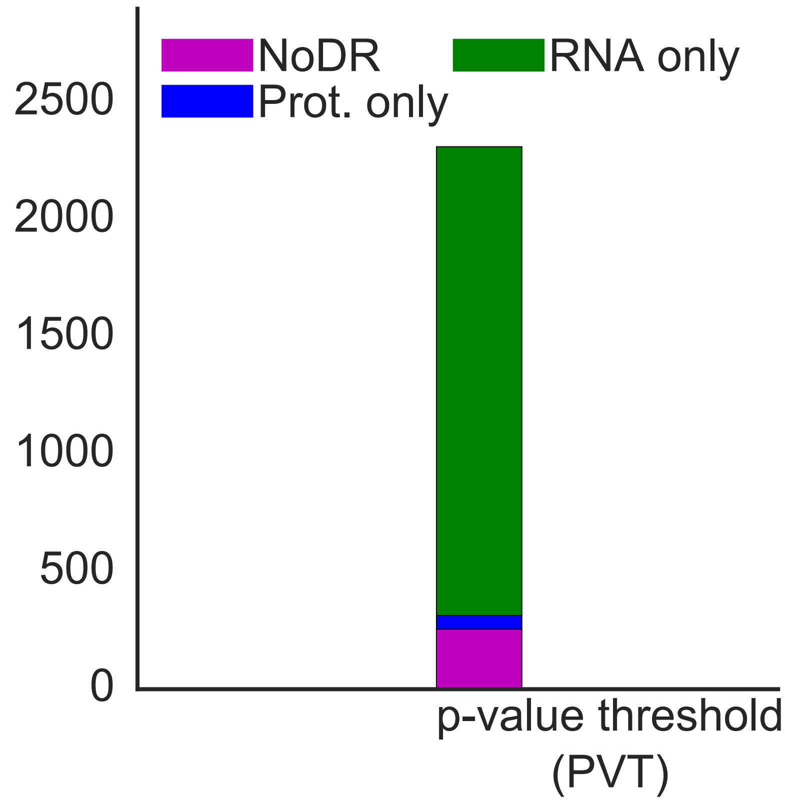




Our method outperforms the p-value threshold method at identifying differential rhythmicity

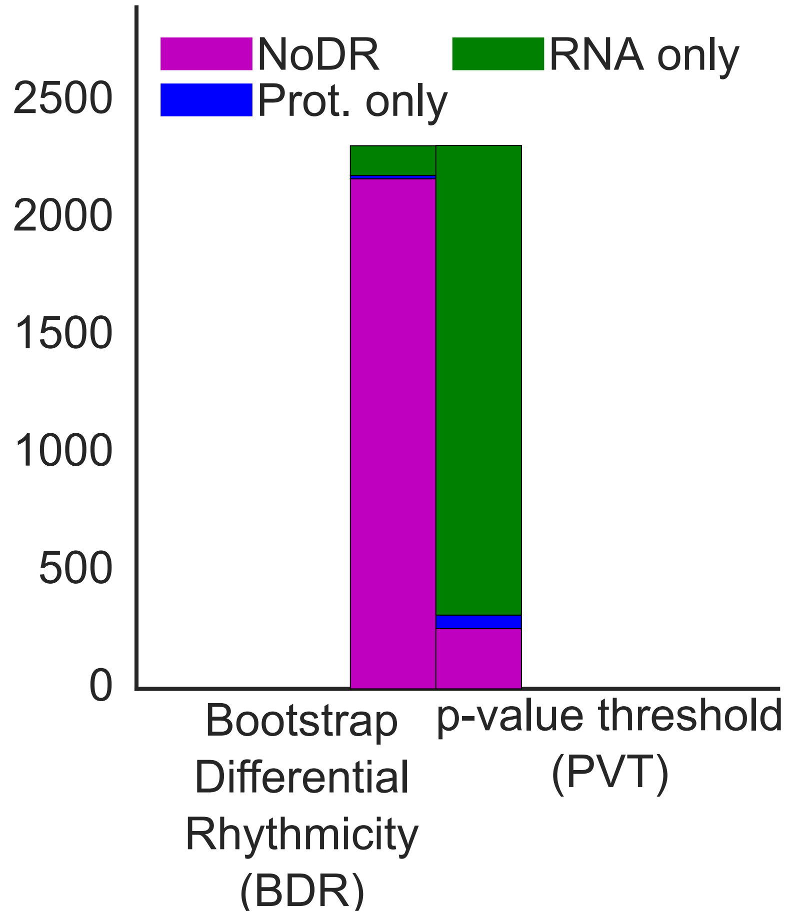


Many studies compare protein level and RNA level rhythmicity



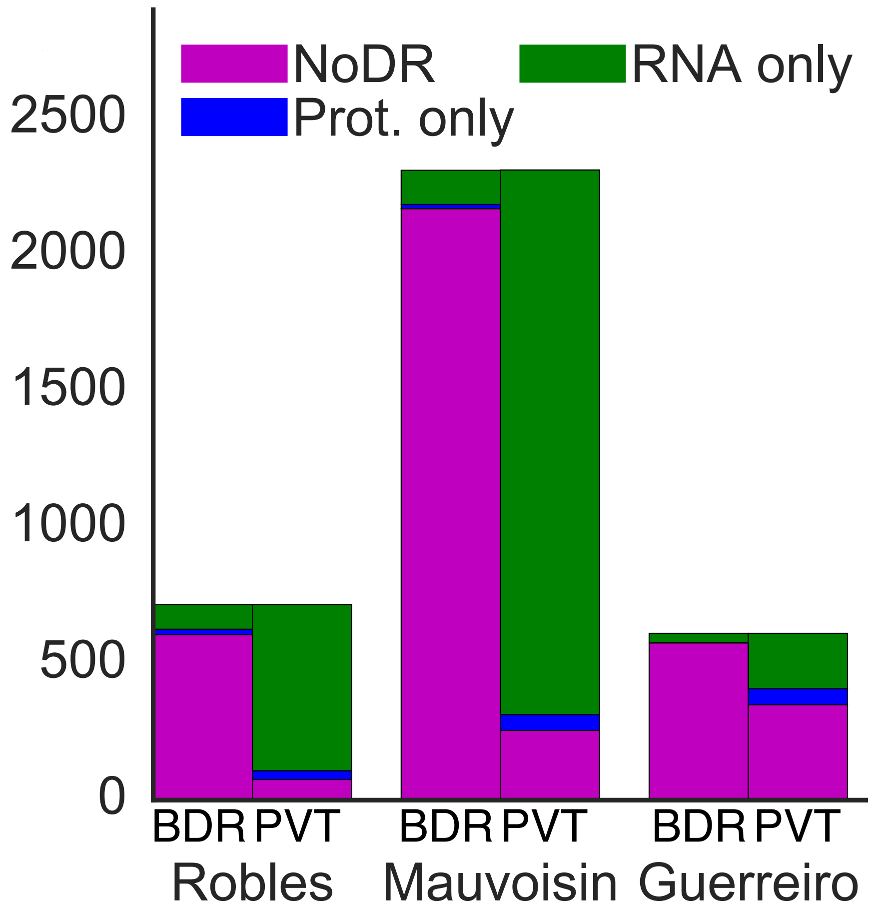
Mauvoisin et al. PNAS 111.1 (2014): 167-172.

We find fewer differences between mRNA and protein time series than the p-value threshold method does



Mauvoisin et al. PNAS 111.1 (2014): 167-172.

We find fewer differences between mRNA and protein time series than the p-value threshold method does



Robles, Maria S., Jürgen Cox, and Matthias Mann. PLoS Genet 10.1 (2014): e1004047.

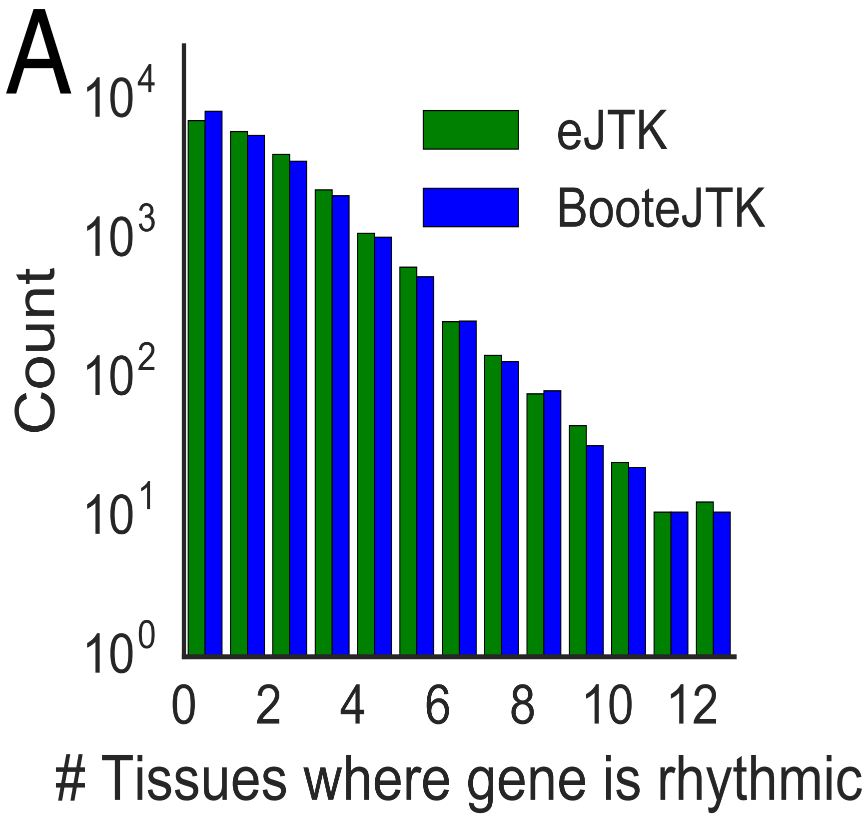
Mauvoisin et al. PNAS 111.1 (2014): 167-172.

Guerreiro et al. Molecular & Cellular Proteomics 13.8 (2014): 2042-2055.

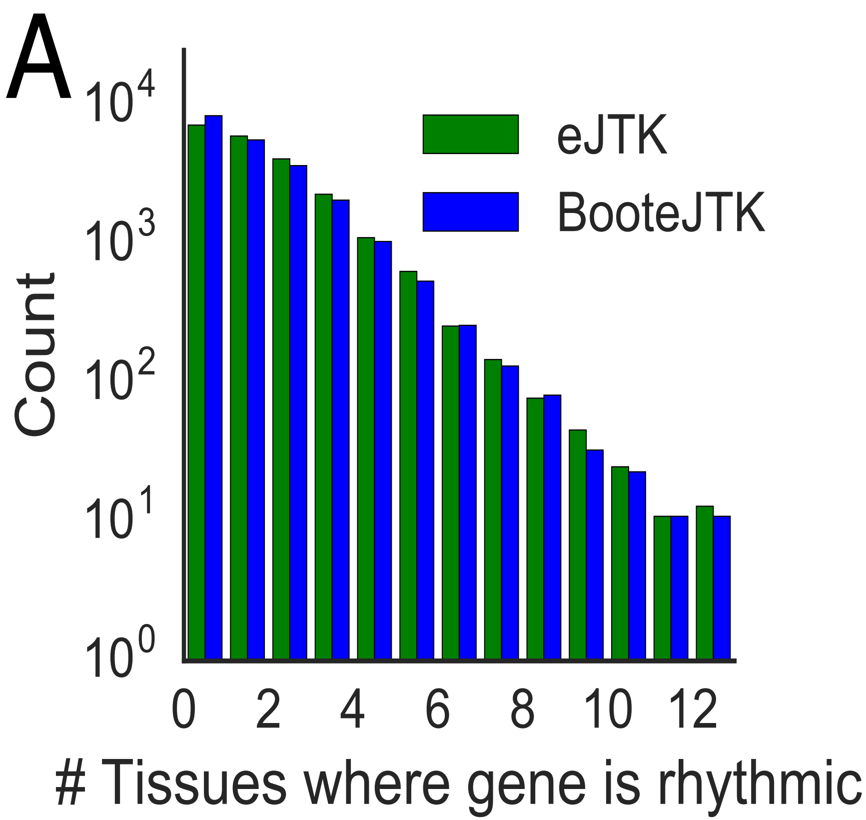
Outline

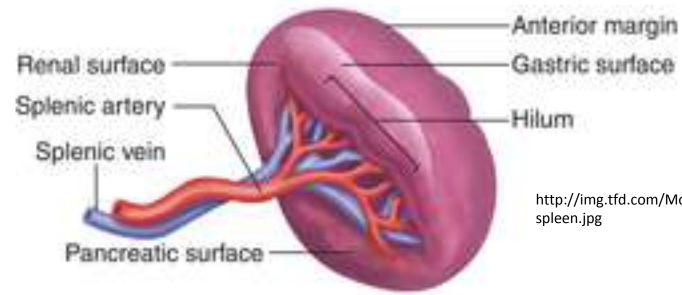
- Biological and Statistical Background
- Improvements to a Rhythm Detection method
- Comparing rhythmicity across conditions
- **Future Directions**

Future direction: Combining information across tissues and conditions



Future direction: Combining information across tissues and conditions





http://img.tfd.com/MosbyMD/thumb/spleen.jpg

Rhythmicity in the spleen?

Questions?

- Biological and Statistical Background
- Improvements to JTK_CYCLE
  - Empirical JTK_CYCLE (eJTK)
    - Searching for asymmetric waveforms
    - Calculating accurate p-values
    - Hutchison *et al.* (2015) "Improved statistical methods enable greater sensitivity for rhythm detection in genome-wide data". *PLoS Computational Biology*. (11) 3
  - Bootstrap eJTK (BooteJTK)
    - Bootstrap resampling time series
      - New to rhythm detection
    - Empirical Bayes variance estimation
      - Common in differential expression analysis
      - New to rhythm detection
    - Greater consistency than eJTK
    - More stringent than eJTK
    - Differences in results are biologically supported
- Comparing rhythmicity across conditions
  - A method that produces accurate p-values for differential rhythmicity
- Future Directions
  - Combining information across conditions and tissues for rhythm detection

Acknowledgements

Dinner Group
Monika Scholz
Jeremy Tempkin
Seyit Kale
Eugene Leypunskiy
Herman Gudjonson
Erik Thiede
Simon Freedman
Catherine Triandafillou
Charlie Matthews
Glen Hocky
Brian Van Koten
Bodhi Vani
Vicky To
Zhiyue Lu

Biophysics 2012
Monika Scholz
Eugene Leypunskiy
Herman Gudjonson
Ryan Mork
Kevin Song

Biophysics Admin.
Dr. Michele Wittels
Dr. Julie Feder
Dr. Adam Hammond
Dr. Tobin Sosnick

PhD Committee
Dr. Aaron R. Dinner
Dr. Mike Rust
Dr. Matthew Stephens

MSTP 2011
Sammy Thomas
Jen Jacobsen
Anya Bershad

MSTP Admin.
Shay McAllister
Dr. Elise Covic
Dr. Alison Anastasio
Dr. Marcus Clark

Collaborators
Dr. Ravi Allada
Dr. Eugene Chang
Dr. Vanessa Leone
Dr. Joseph Bass

DARPA

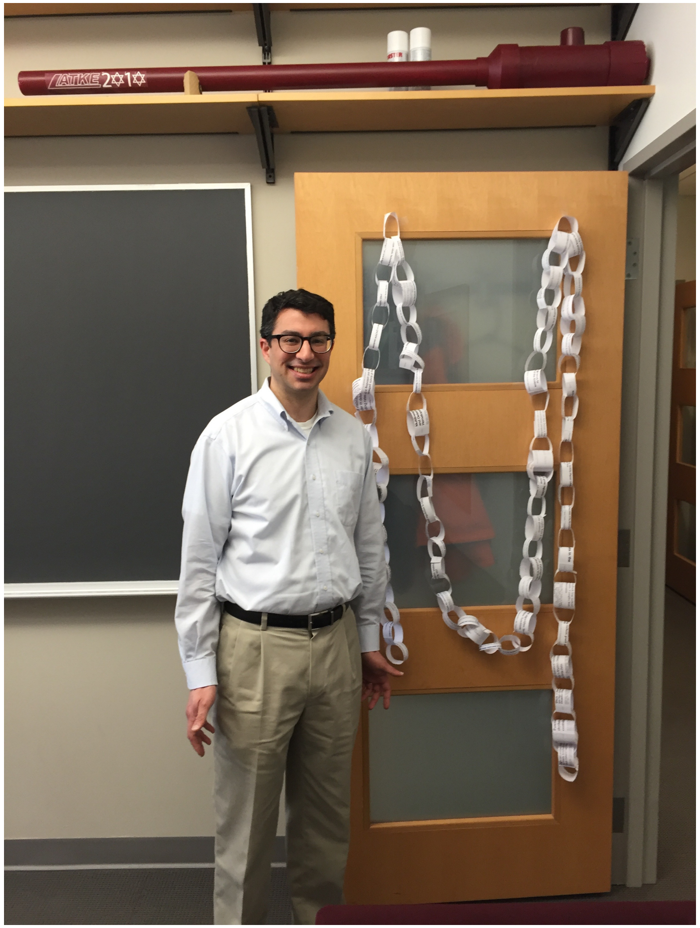
NATIONAL INSTITUTES OF HEALTH

Kate Carter

My parents, Barb and Rick

Monika Scholz

Aaron Dinner



Acknowledgements

Dinner Group
Monika Scholz
Jeremy Tempkin
Seyit Kale
Eugene Leypunskiy
Herman Gudjonson
Erik Thiede
Simon Freedman
Catherine Triandafillou
Charlie Matthews
Glen Hocky
Brian Van Koten
Bodhi Vani
Vicky To
Zhiyue Lu

Biophysics 2012
Monika Scholz
Eugene Leypunskiy
Herman Gudjonson
Ryan Mork
Kevin Song

MSTP 2011
Sammy Thomas
Jen Jacobsen
Anya Bershad

Biophysics Admin.
Dr. Michele Wittels
Dr. Julie Feder
Dr. Adam Hammond
Dr. Tobin Sosnick

MSTP Admin.
Shay McAllister
Dr. Elise Covic
Dr. Alison Anastasio
Dr. Marcus Clark

PhD Committee
Dr. Aaron R. Dinner
Dr. Mike Rust
Dr. Matthew Stephens

Collaborators
Dr. Ravi Allada
Dr. Eugene Chang
Dr. Vanessa Leone
Dr. Joseph Bass