Causality Assessment with Multiple Time Series Data

Curtis A. Bagne, Ph.D. – DataSpeaks, Inc.

Brian D. Athey, Ph.D., Chair, Department of Computational Medicine and Bioinformatics, University of Michigan

Edward Barbour, The Rockefeller University

Walter Meixner, U-M

Alex Ade, U-M

Intent

- DataSpeaks, Inc., an early stage growth company, offers a *uniquely digital measurement-by-computation tool* embodied in software
- It offers a disruptive measurement science solution
- Collaborate (research, grants, publications, etc.) to:
  - Help you – DCM&B, U-M, etc.
  - Advance DataSpeaks
  - Help solve the (i) bench to bedside and (ii) clinical science to clinical practice translation problems
  - Advance P4 Medicine

Causality Assessment

- Central to scientific understanding
- Focus on Complex Dynamic and Adaptive Systems (CDAS)
- Essential for basic and applied science (e.g., medicine)
- Two contexts for DataSpeaks' computational method
  1. Without randomized experimental control
     - HeLa cell cycle control – a human cancer cell line since 1951
     - Cancer – a failure of cell cycle control
  2. With randomized experimental control
     - Frequentist and Bayesian statistics not sufficient
     - Need Ultra RCTs (Randomized Controlled Trials) with DataSpeaks and statistics

Time Series Data

- Here a shorthand for "periodic time-ordered data"
- Inclusive of:
  - Repeated measurements (< 20 repeats)
  - Time series (20 or more repeats)
  - Two or more repeats, hopefully many more
- Contrasts with cross-sectional data (including change scores)
- Time series can provide orders of magnitude more information to *understand individuals scientifically*
- Periodicity helpful for understanding temporal dynamics

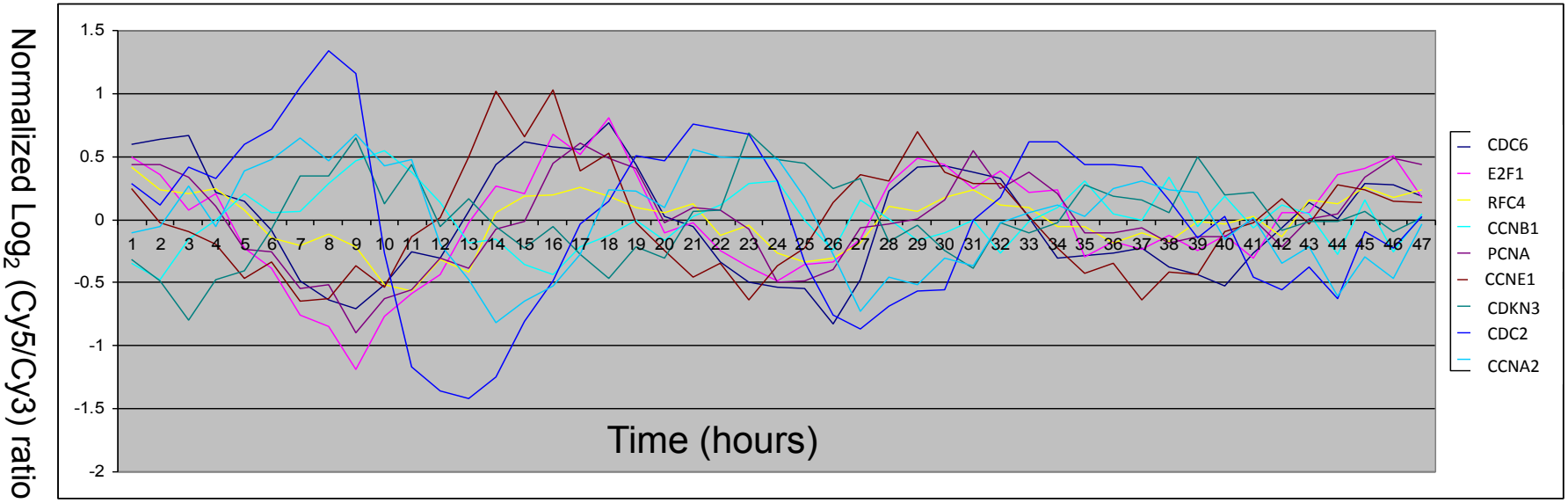
The Great Bottleneck: We Need Better Computational Tools

Inputs:
- Omic sciences
- Data collection technologies
- IT infrastructure
- EHR

Outputs:
- P4 Medicine
- Safer medicine
- Better and more affordable health and healthcare
- Better drugs to market faster
- New drug indications

What problem could not be done better with multiple time series?

HeLa Cell Data



- 47 repeated hourly measurements
- About three cell cycles
- Synchrony deteriorates over time
- Goal: Elucidate temporal causal networks

HeLa Cell Data: Source

BIOINFORMATICS

Vol. 26 ECCB 2010, pages i517–i523
doi:10.1093/bioinformatics/btq377

Discovering graphical Granger causality using the truncating lasso penalty

Ali Shojaie* and George Michailidis
Department of Statistics, University of Michigan, Ann Arbor Michigan 48109, USA

We acknowledge George Michailidis for providing these HeLa cell data.

Granger Causality

- If “a signal X_1 "Granger-causes" (or "G-causes") a signal X_2, then past values of X_1 should contain information that helps predict X_2 above and beyond the information contained in past values of X_2 alone.”
- “Its mathematical formulation is based on linear regression modeling of stochastic processes.”
- This won Clive Granger the 2003 Nobel Prize in Economics
- One application of DataSpeaks is a *measurement* alternative to Granger causality, not a variation

HeLa Cell Data: History

Molecular Biology of the Cell
Vol. 13, 1977–2000, June 2002

Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors[D]

Michael L. Whitfield,* Gavin Sherlock,* Alok J. Saldanha,* John I. Murray,* Catherine A. Ball,* Karen E. Alexander,† John C. Matese,* Charles M. Perou,‡ Myra M. Hurt,† Patrick O. Brown,§‖¶ and David Botstein*¶

- Whitfield, M. *et al* uses clustering methods
- Same HeLa cell data subject of other publications with various methods in addition to Shojaie/Michailidis
  - Lozano *et al* - 2009 (Grouped Graphical Granger Modeling)
  - Sambo *et al* – 2008 (CNET)
  - Sacchi *et al* – 2007 (Precedence Temporal Networks)

DataSpeaks

- Offers uniquely digital measurement-by-computation algorithm embodied in software
- MQALA (Method for the Quantitative Analysis of Longitudinal Associations)
- Applies to multiple time series data
- Measures the (i) amount, (ii) positive or negative direction and (iii) strength of evidence for coordination of action

DataSpeaks, continued

- Coordination scores (CS) describe and help predict how *individual* CDAS *work* over time
  - Function internally (e.g., HeLa cell cycle control)
  - Respond to environments including treatments (e.g., Ultra RCTs)
  - Act as agents on their environments (neglected stepchild of science)
- Bagne holds two issued U.S. software patents
  - 6,317,700 – Method and System to Perform Empirical Induction
  - 6,516,288 – Method and System to Construct Action Coordination Profiles
- Software available as a functional prototype, open to additional collaboration and project discussions
- Dennis Nash heads business development dnash@dataspeaks.com
- Collaborating with U-M through Brian Athey

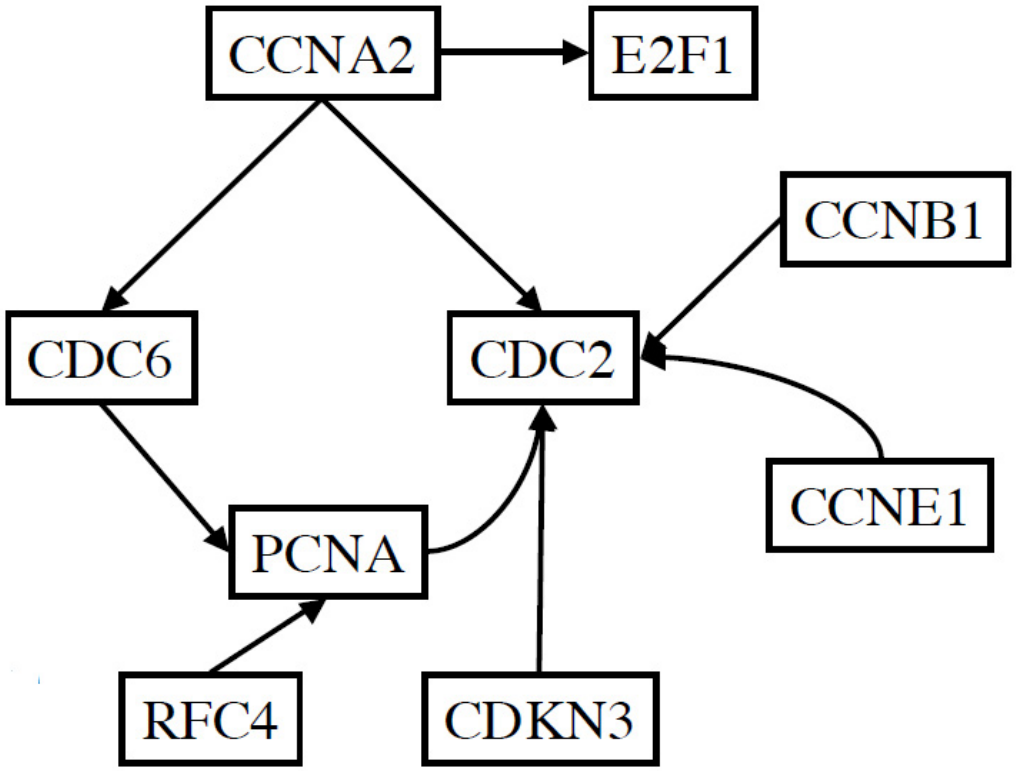
DataSpeaks' Approach: Causality Assessment with HeLa Cell Cycle Control Data

- Nine genes, cancer activators
- 72 pair-wise directional interactions over time

  (e.g., CDC6➔CDC2, PCNA➔E2F1) [➔ operates on]

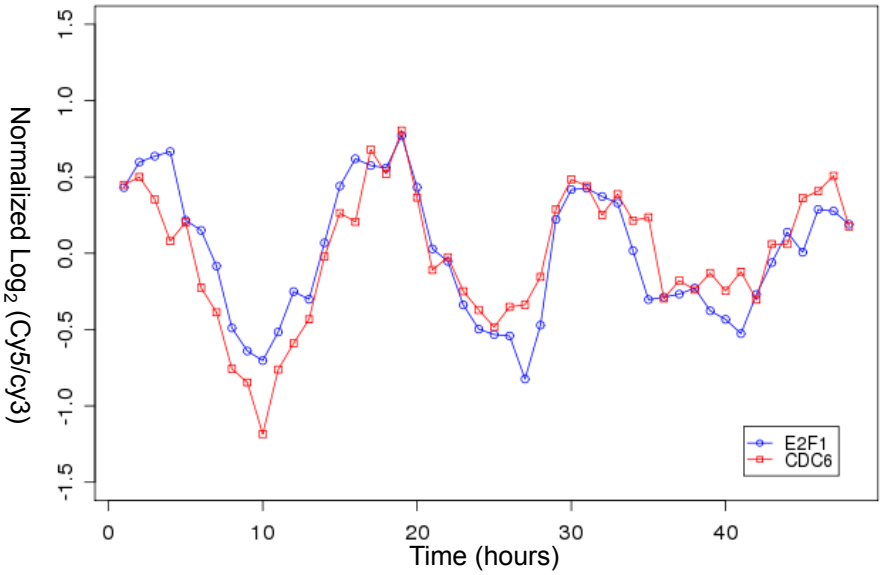
  [independent variable (IV) ➔ dependent variable (DV)]
- 36 pairs (e.g., PCNA➔CDC2, CDC2➔PCNA; CDKN3➔CDC2, CDC2➔CDKN3)
- *Measure temporal asymmetries* within each pair using DataSpeaks' Coordination Scores (CS)
- Causes must come before effects – the temporal criterion of causal relationships

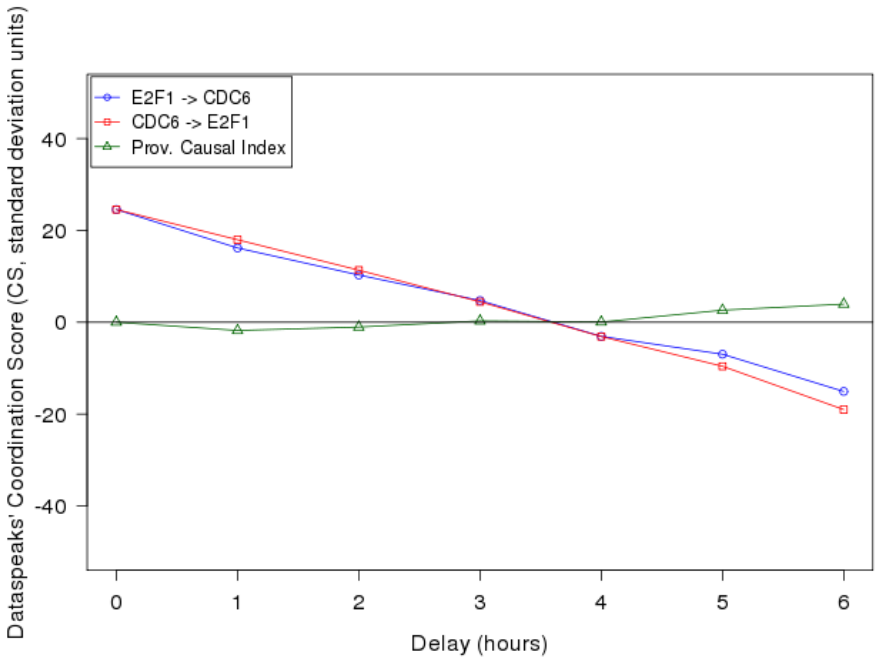
“Known” Network



- “Known” network from Sambo, et al 2008
- Extracted from www.thebiogrid.org
- Cited by Shojaie/Michailidis as “Known Regulatory Network”

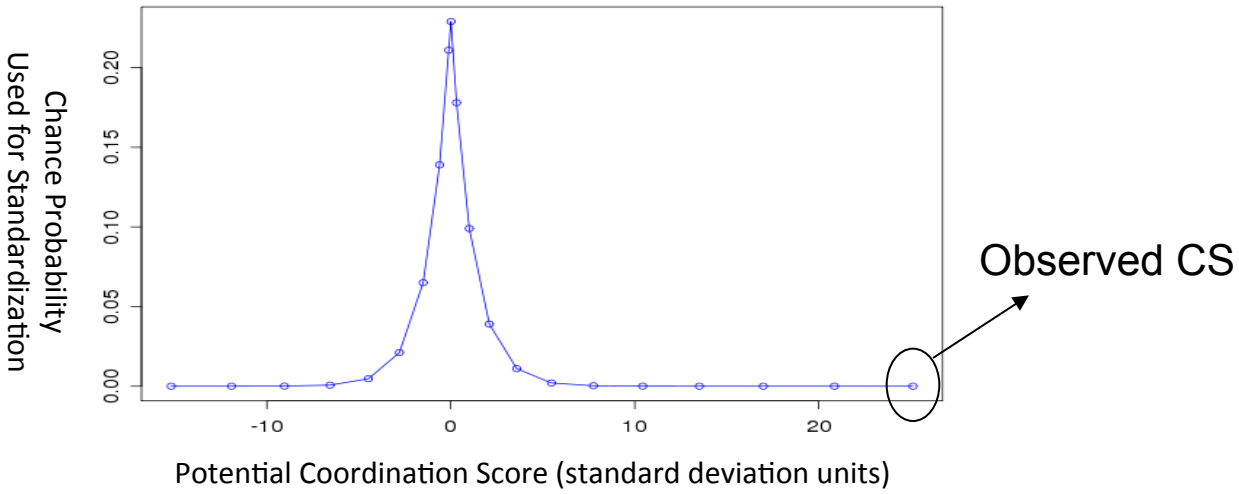
E2F1➔CDC6, CDC6➔E2F1 Results





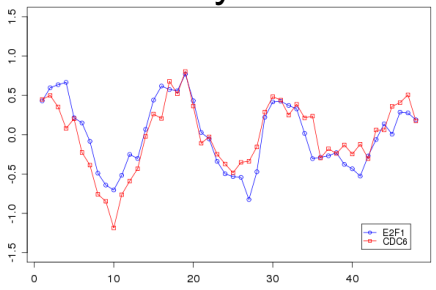
Meaning of Coordination Scores

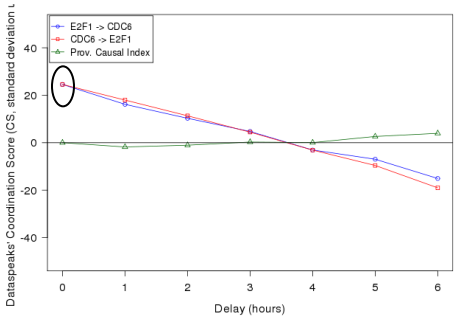
- Coordination - “harmonious functioning of parts for effective results”
- The CS at Delay = 0 in the previous slide is 25.123
- CSs are in standard deviation units
- 25.123 is one score in the following distribution of potential scores



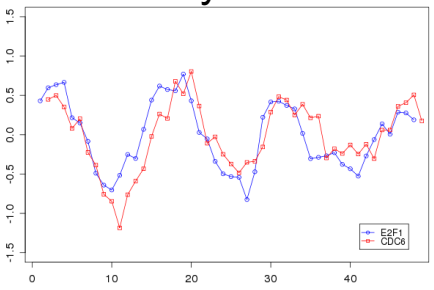
- We will introduce you to computing these scores after several more results

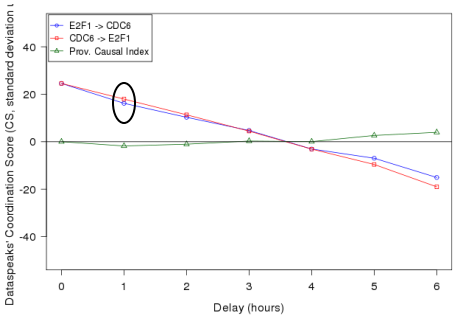
Delay = 0



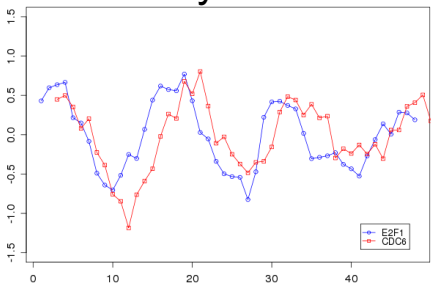


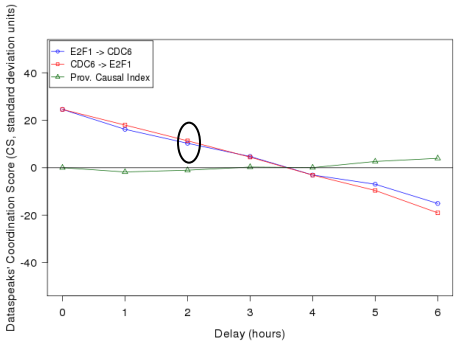
Delay = 1



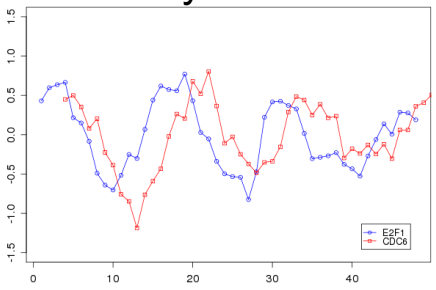


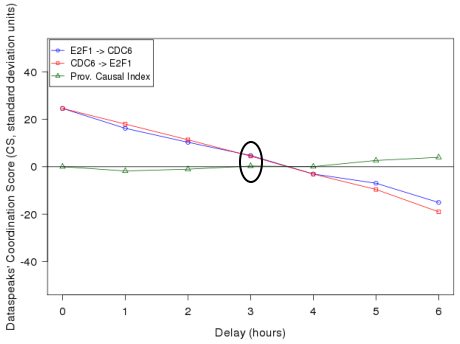
Delay = 2



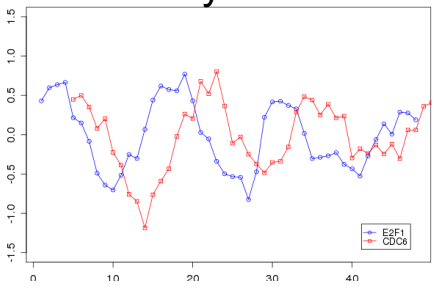


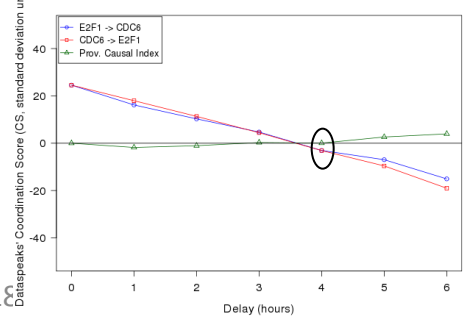
Delay = 3



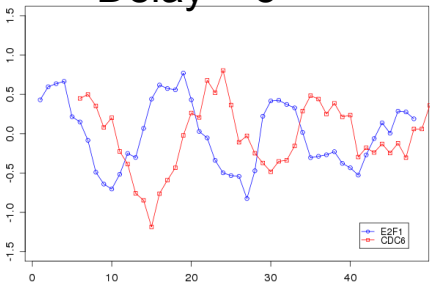


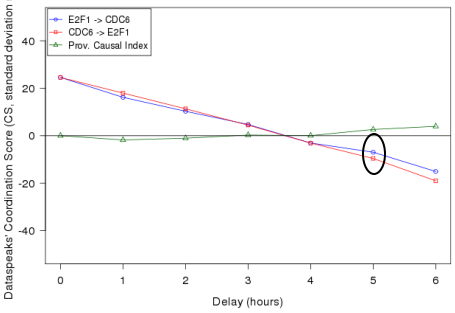
Delay = 4



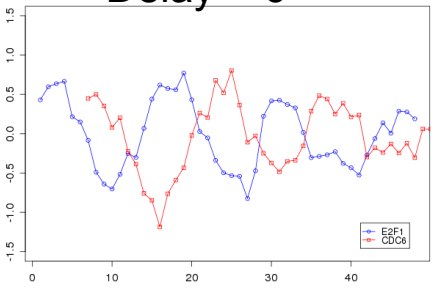


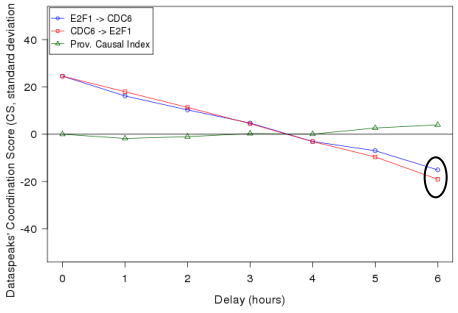
Delay = 5



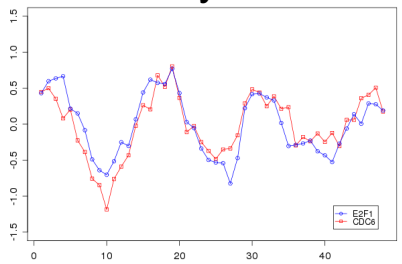


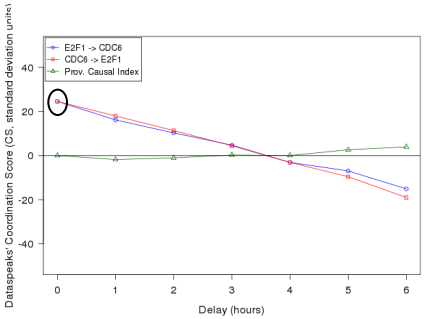
Delay = 6



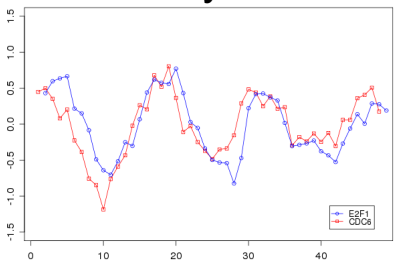


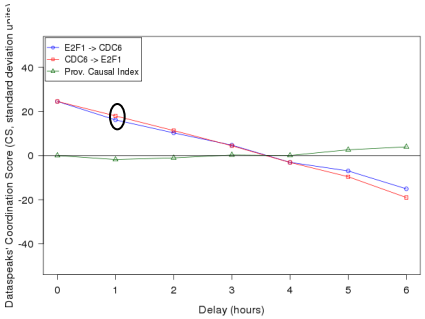
Delay = 0



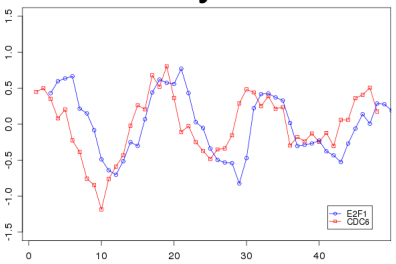


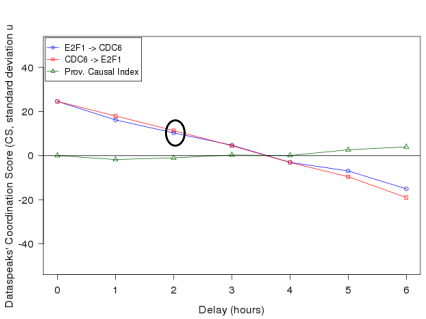
Delay = 1



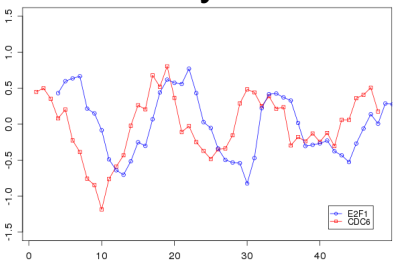


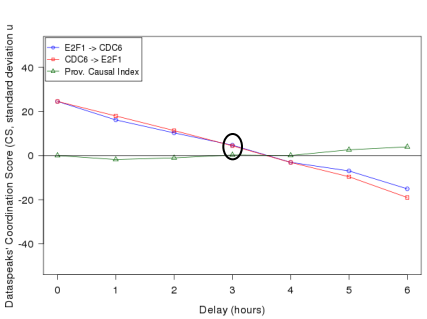
Delay = 2



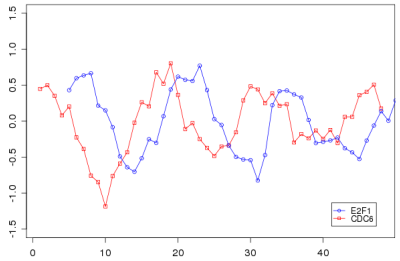


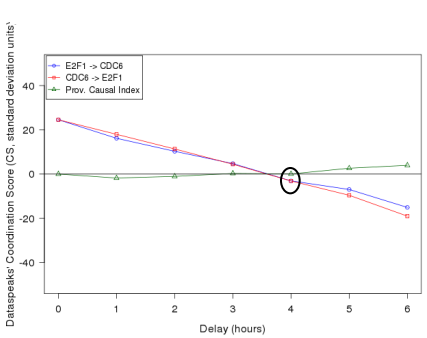
Delay = 3



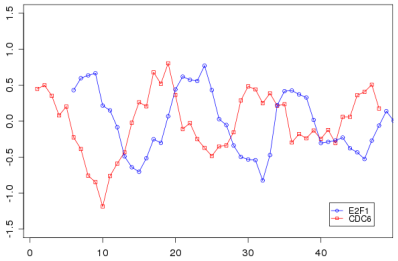


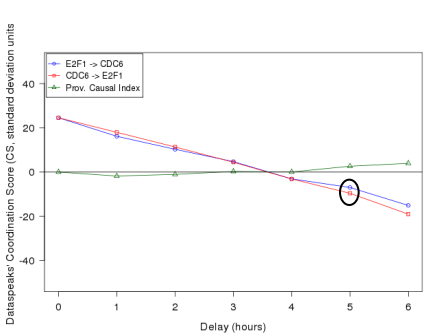
Delay = 4



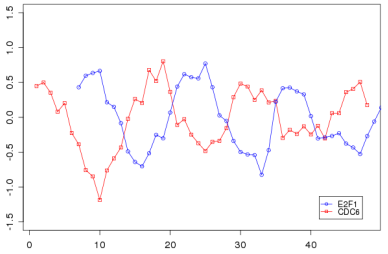


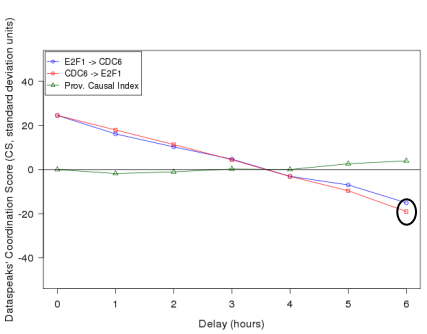
Delay = 5



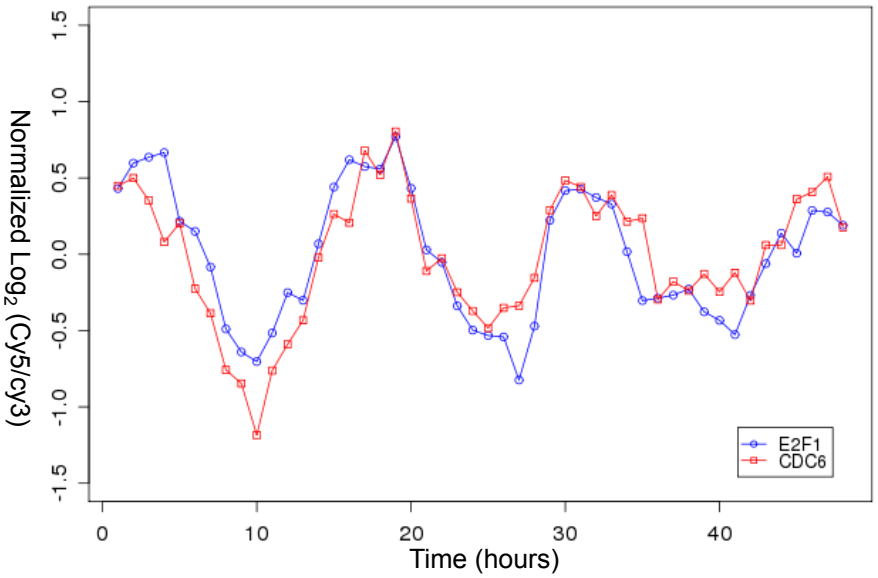


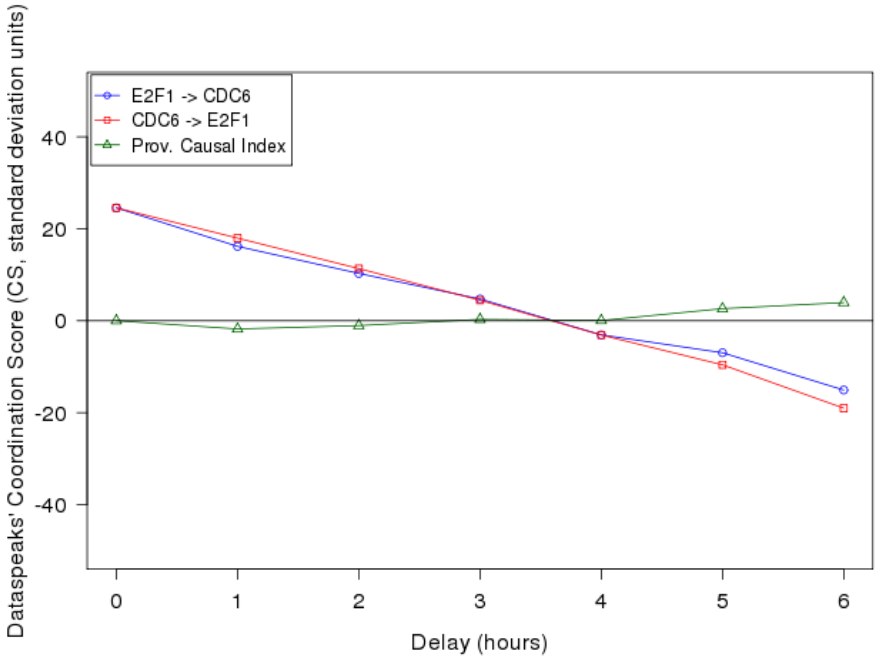
Delay = 6





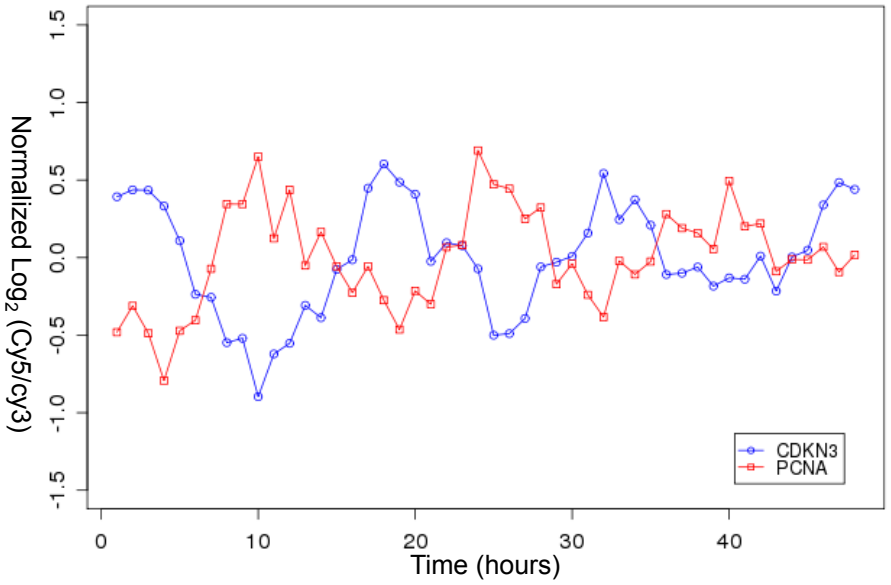
E2F1➔CDC6, CDC6➔E2F1 Results

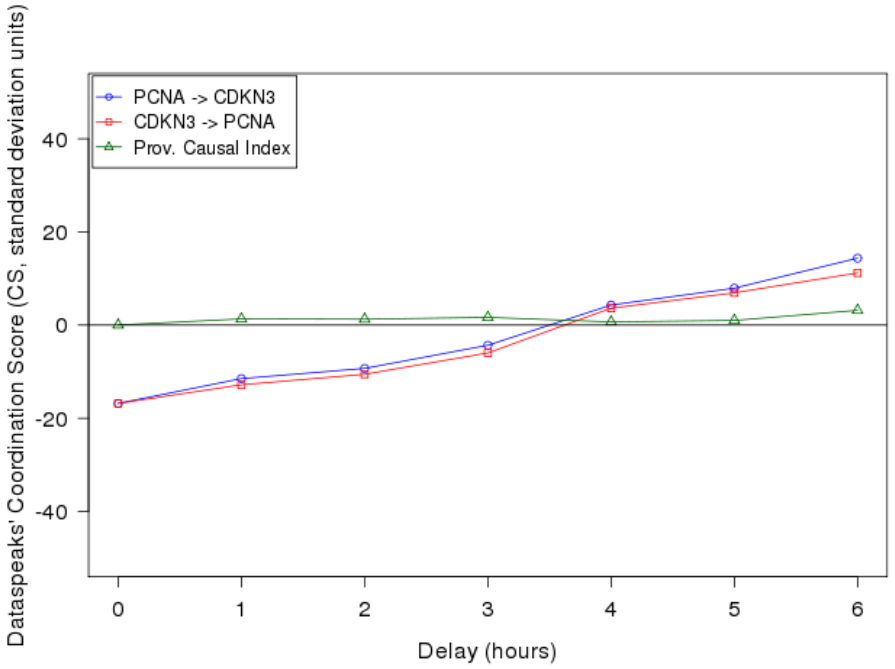




- Not a “known” interaction
- Apt to be clustered together but not causal
- DataSpeaks - almost no evidence of causality (green line)
- Not identified by Shojaie/Michailidis

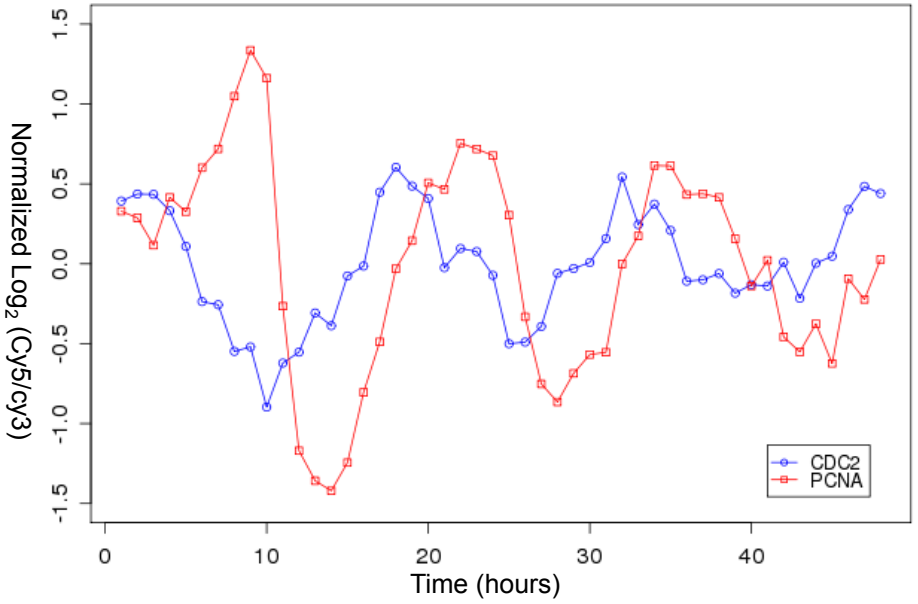
PCNA➔CDKN3, CDKN3➔PCNA Results

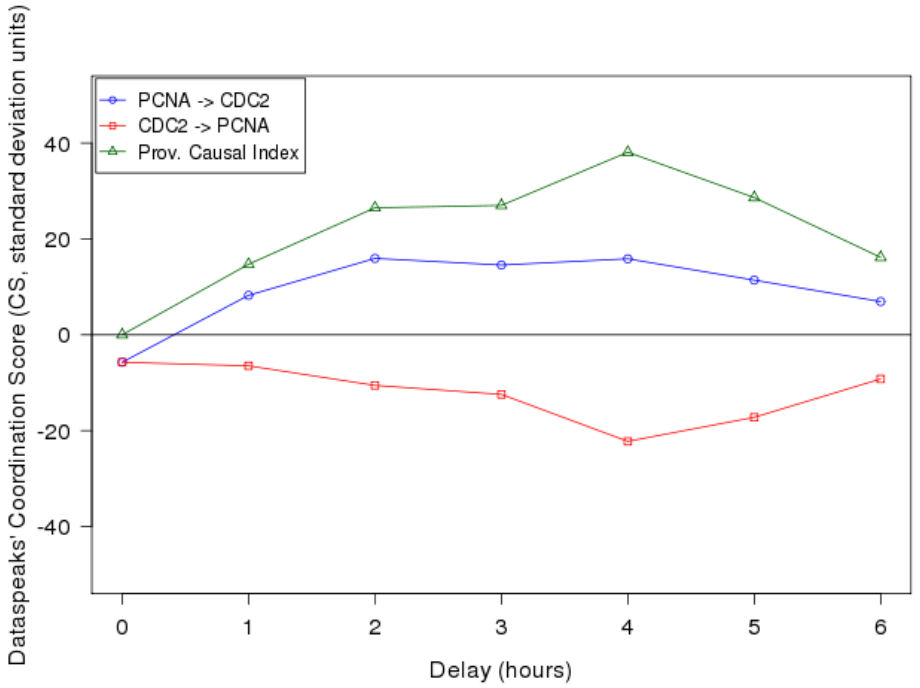




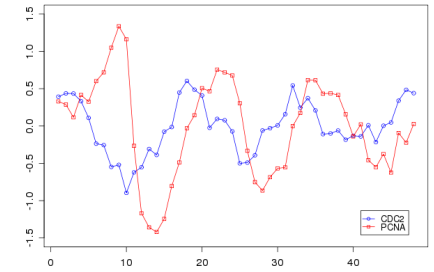
- Not a “known” interaction
- DataSpeaks - almost no evidence of causality (green line)
- Not identified by Shojaie/Michailidis

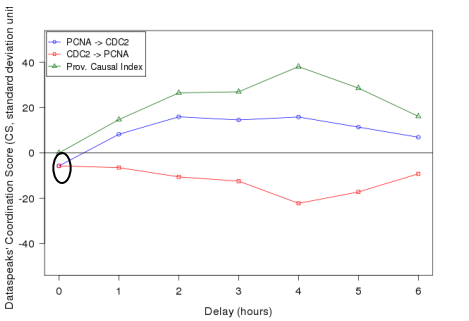
PCNA➔CDC2, CDC2➔PCNA Results



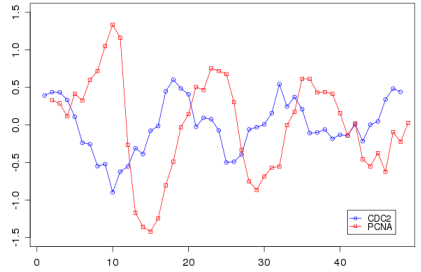


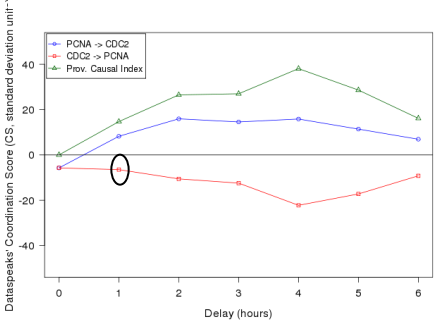
Delay = 0



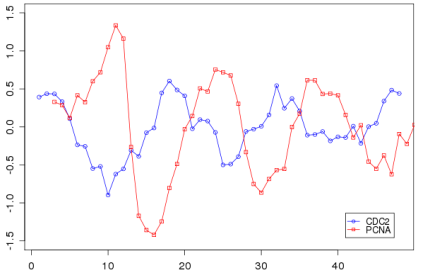


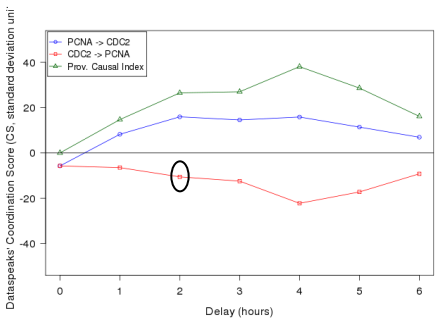
Delay = 1



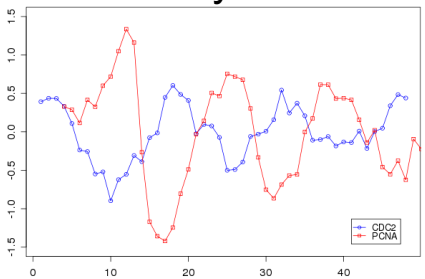


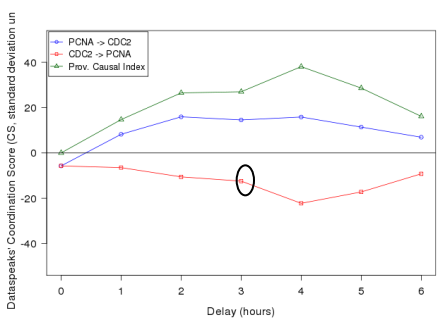
Delay = 2



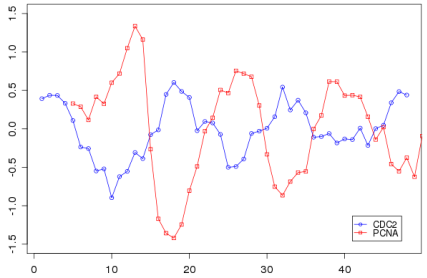


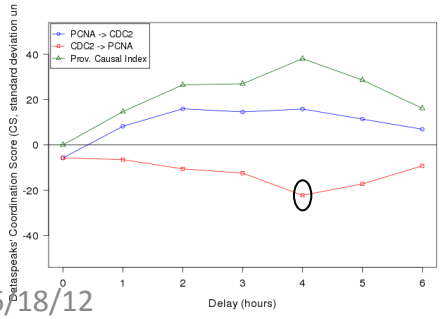
Delay = 3



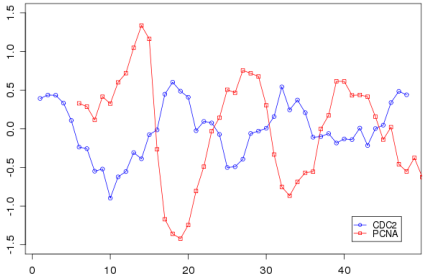


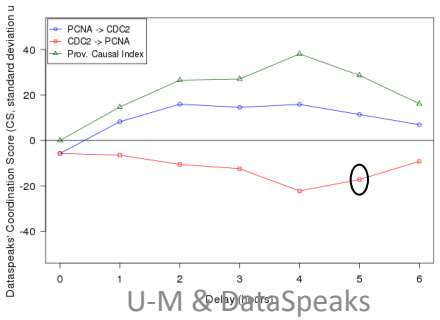
Delay = 4



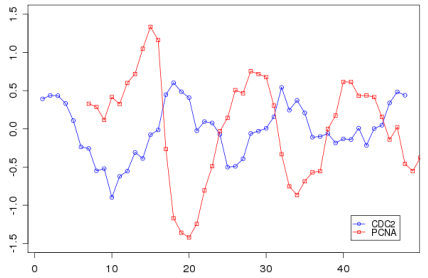


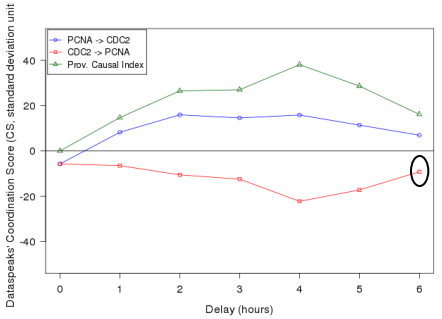
Delay = 5



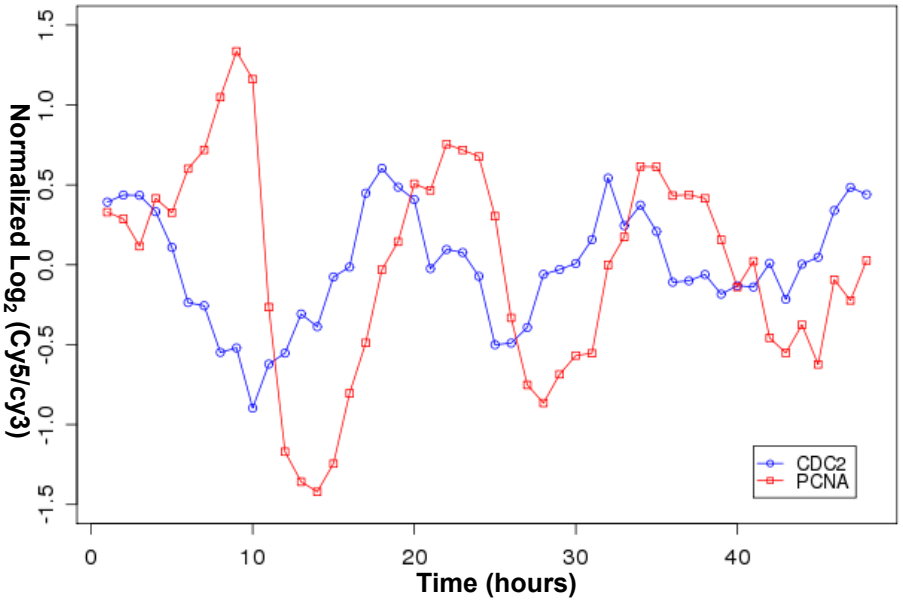


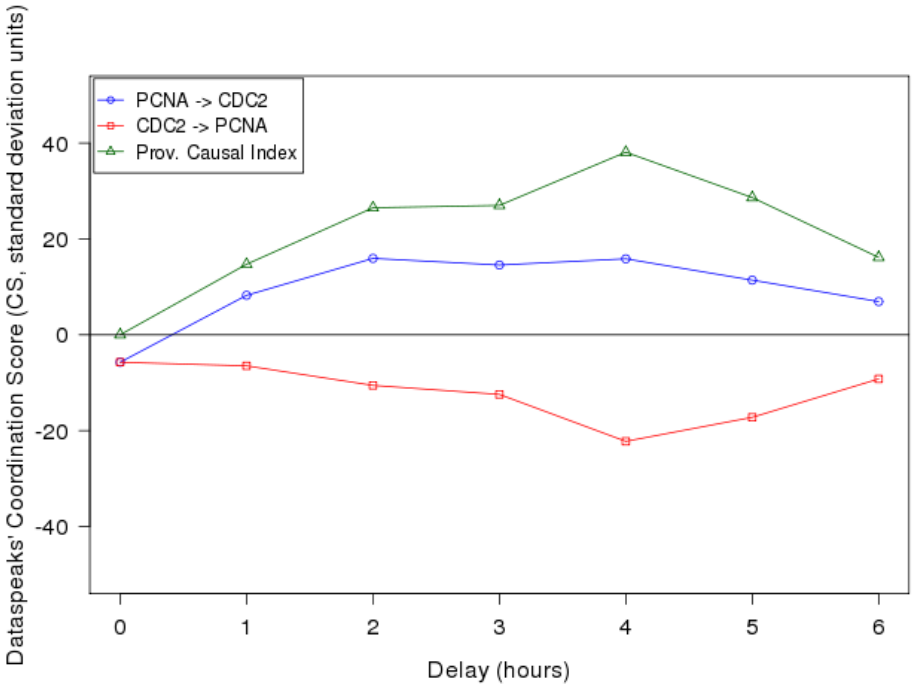
Delay = 6





PCNA➔CDC2, CDC2➔PCNA Results



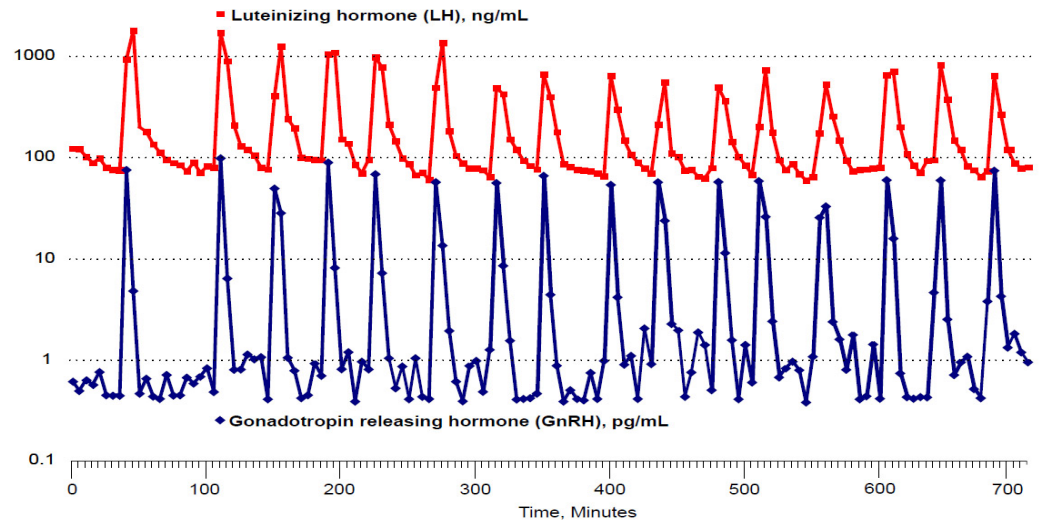


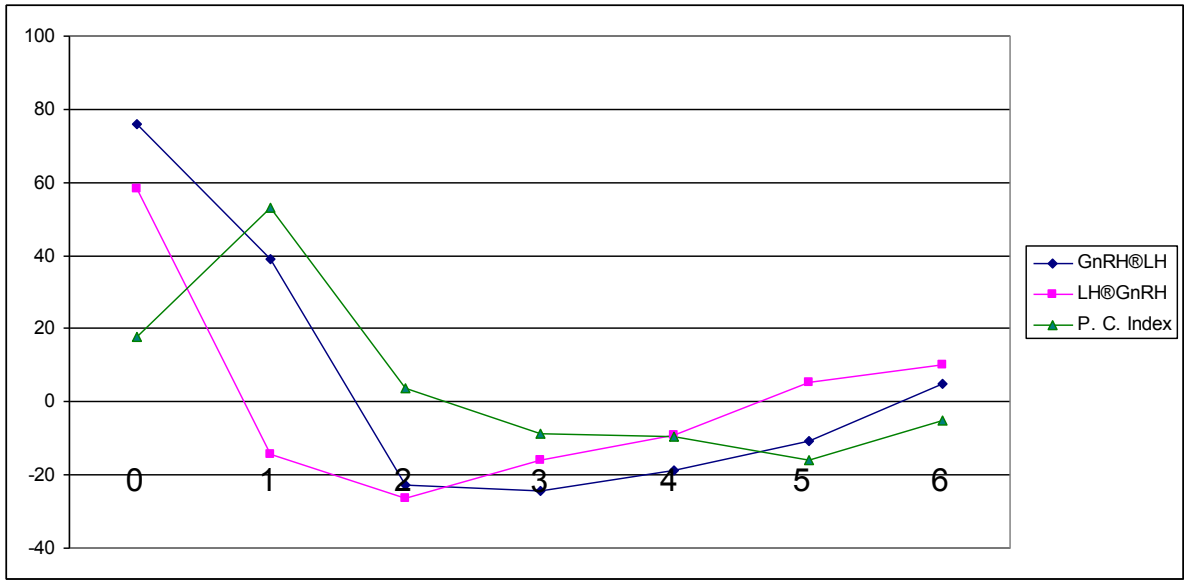
- This is a “known” interaction
- DataSpeaks identifies strong evidence for causality
- Not detected by Granger causality as reported by Shojaie/Michailidis

Results Summary: Comparison by Method

- This is summarized in a handout

Hormone Data Example





Major Steps to Compute and Use DataSpeaks' CS Scores

1. Optional data pre-processing
2. Convert each dimensional series into a set of digital series
3. Form additional digital series to account for delay, persistence, episodes, Boolean events
4. Cross-classify each digital series for an IV or predictor variable with each digital series for for an DV or predicted variable to form arrays of 2x2 tables
5. Compute a raw CS for each 2x2 table
6. Compute a standardized CS for each 2x2 table
7. Summarize standardized CS scores
8. Analyze results statistically when there are groups of two or more individuals

Step 2: Digitization

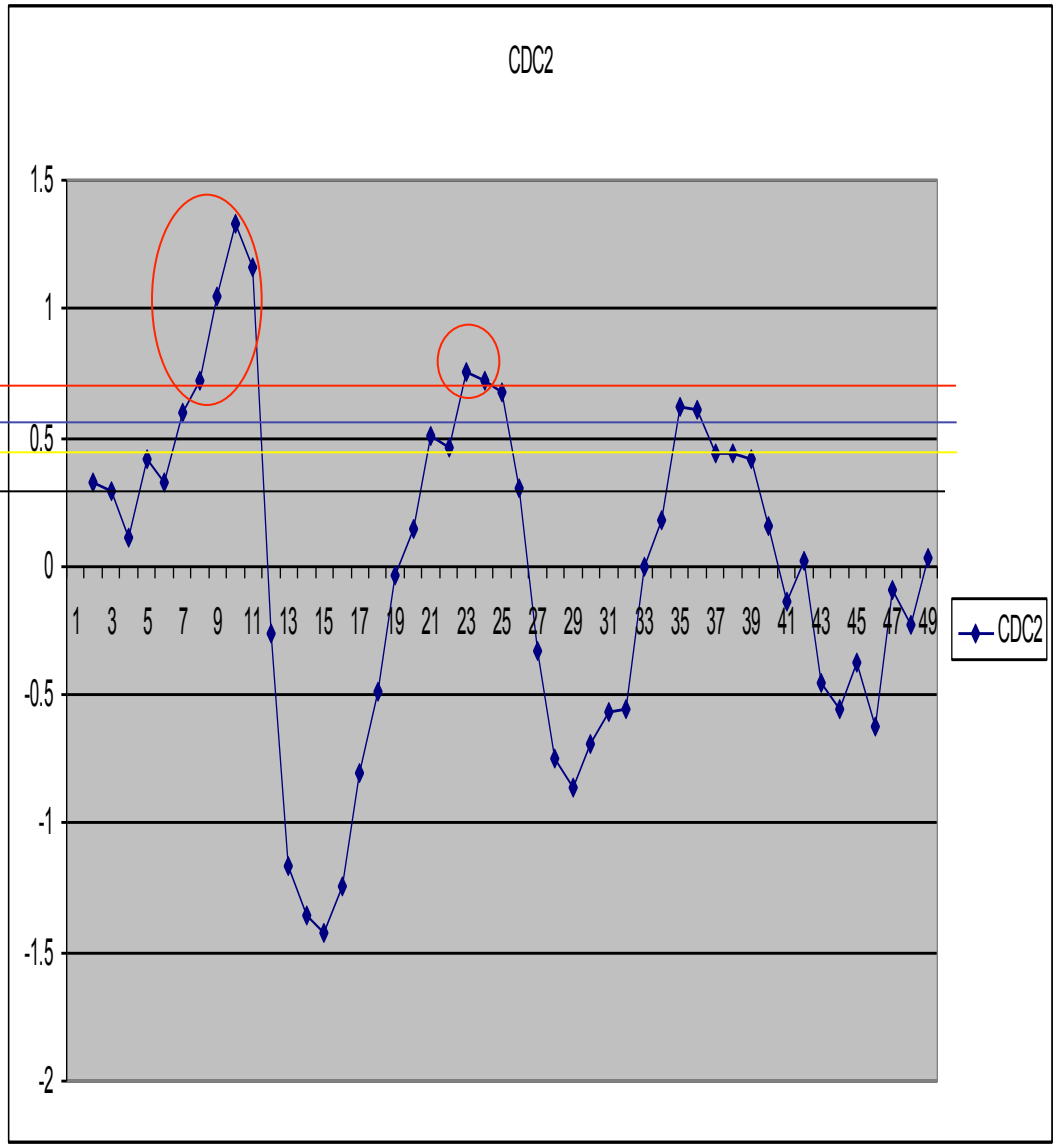
- One key to DataSpeaks' capabilities
- Required when a time series has more than two levels
- *Digitization has potential to be as valuable for understanding individuals scientifically with time series as digitization has been for photography and communications.*
- Individuals can be cells, brains, people, whole populations, economies, Earth's biosphere, etc.

Digitization Summary for CDC2

"Action Digigram"

Dimensional Resolution

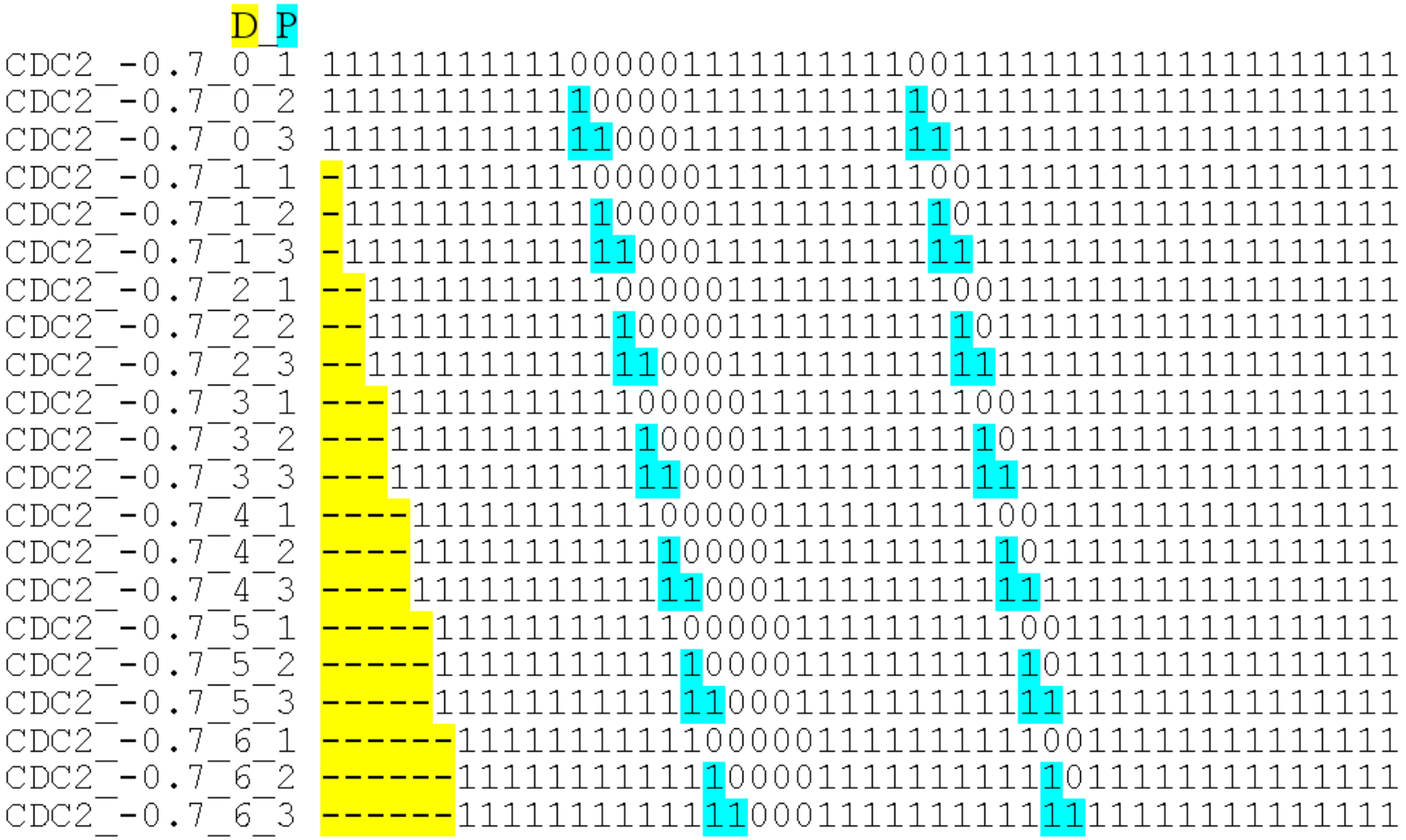
CutPoints	Temporal Resolution
0.714	00000011110000000000011000000000000000000000000000
0.584	00000111110000000000011100000000011000000000000000
0.453	00000111110000000000111110000000001100000000000000
0.322	10011111110000000001111100000000011111000000000000
0.192	11011111110000000001111110000000011111000000000000
0.061	11111111110000000011111111000000011111110000000000
-0.070	11111111110000000111111110000000111111101000000 1
-0.200	11111111110000000111111110000000111111111110000101
-0.331	11111111110000000111111110000000111111111110000111
-0.461	11111111110000000111111111000000111111111111010111
-0.592	11111111110000001111111111000111111111111111110111
-0.723	11111111110000001111111111001111111111111111111111



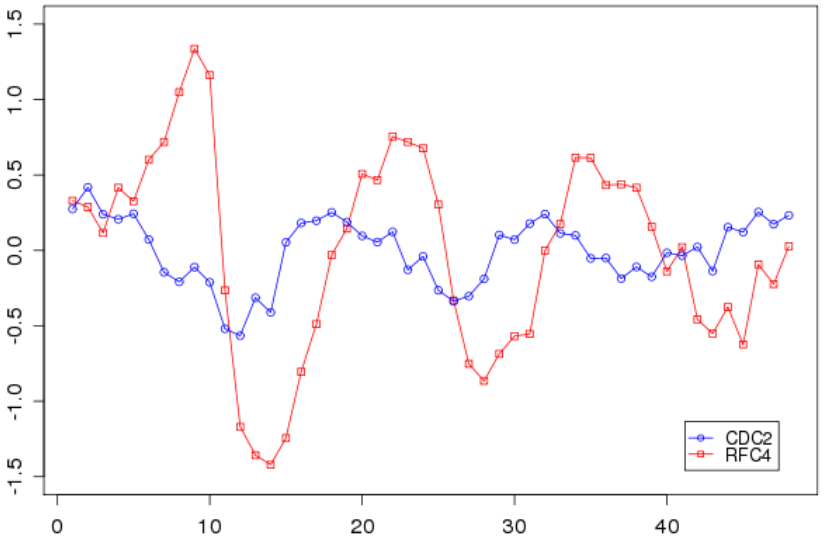
Value of High Resolution – Temporal and Dimensional

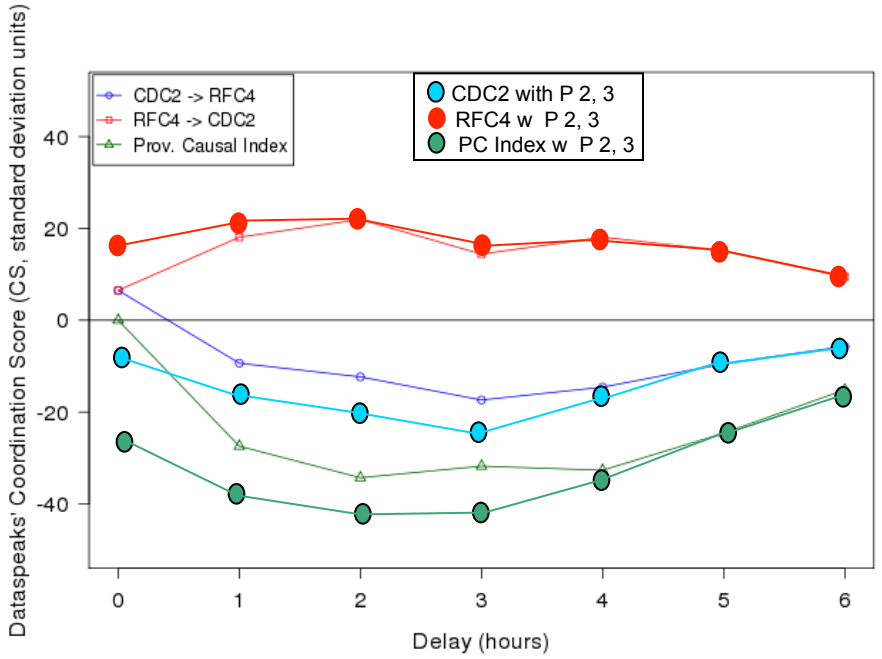
- Better resolution of DataSpeaks for understanding *dynamic temporal phenomena* is related to better resolution in digital photography.
- An aspect of *DataSpeaks' data microscope* – see patterns that have never been seen before
- Biggest gains apt to come from improving temporal resolution.
- *Statistics gains power with more subjects.*
- *DataSpeaks gains power with more repeated measurements.*
- **Big Data** – Really gain power with more subjects AND more repeated measurements AND higher resolution.

Step 3: Form Additional Digital Series – e.g., Delay (D) and Persistence (P)



RFC4➔CDC2, CDC2➔RFC4
Delay and Persistence





Capabilities to Define Digital Events: DataSpeaks' Software Currently

- Independent events
  - 12 levels of dimensional resolution
  - 7 levels of Delay, 0 – 6
  - 5 levels of Persistence, 1 – 5
  - 36 combinations of Episode Length and Episode Criterion
  - 15,120 (12x7x5x36) total combinations
- Dependent events
  - 12 levels of dimensional resolution
  - 36 combinations of Episode Length and Episode Criterion
  - 432 (12x36) total combinations
- Superb for pattern finding

Boolean Independent Events for Complexity

- Assess *multiple causes*, e.g. multiple levels of:
  - Gene activity, proteins, lipids, carbohydrates, metabolites
  - Electrophysiological variables, brain activity
  - Drugs – an alternative to usual way of investigating drug/ drug interactions
- Also can do Boolean dependent events for syndromes
  - Multiple signs and symptoms of disease and disorder
- An antidote for reductionism?

Example: Boolean Independent Events

- PCNA_0.2_1_1_0_1
  111110000000000111100000000001110000000000111
- CCNA2_0.1_1_1_0_1
  000001111111000000110111110000001111111010000 01
- PCNA_0.2_1_1_0_1 AND CCNA2_0.1_1_1_0_1
  000000000000000001100000000000010000000000001
- PCNA_0.2_1_1_0_1 OR CCNA2_0.1_1_1_0_1
  111111111111000011110111110000011111111010001 11

Step 4: Cross-Classify Digital Series to Form an Array of 2x2 Tables

- Cross-classify each digital series for an IV with each digital series for a DV

			IV			
			Present	Absent		
			1	0		
	DV	Present, 1	1, 1 $a = 23$	0, 1 $b = 4$	27	
		Absent, 0	1, 0 $c = 17$	0, 0 $d = 4$	21	
			40	8		

- Example:
- PCNA_-0.4_1_1_0_1 1111111000001111111111110001111111111111111111
- CDC2_0.1_1_1 1101111111000000001111111000000111111101000010 1
- *aacaaaabbbddccccccaaaaaabddccccaaaaaaaacacccccaca*
- Yields 3-D array of 12x12x7=1008 2x2 tables for each gene pair for these examples

Step 5: Compute Raw Coordination Scores

- Start with observed 2x2 table
- Compute chi square
- Set sign
  - Compute expected value of *a*, E(*a*)
  - If observed value of *a*, O(*a*), is < E(*a*), then raw CS is negative chi square to indicate negative CS
  - If O(*a*) is > E(*a*), then raw CS is positive chi square to indicate positive CS

Step 6: Compute Standardized Coordination Scores

- CSs must be standardized
- DataSpeaks standardizes these scores to have an expected value of 0 and a standard deviation of 1.
- How? What is the trick?

Standardization Example

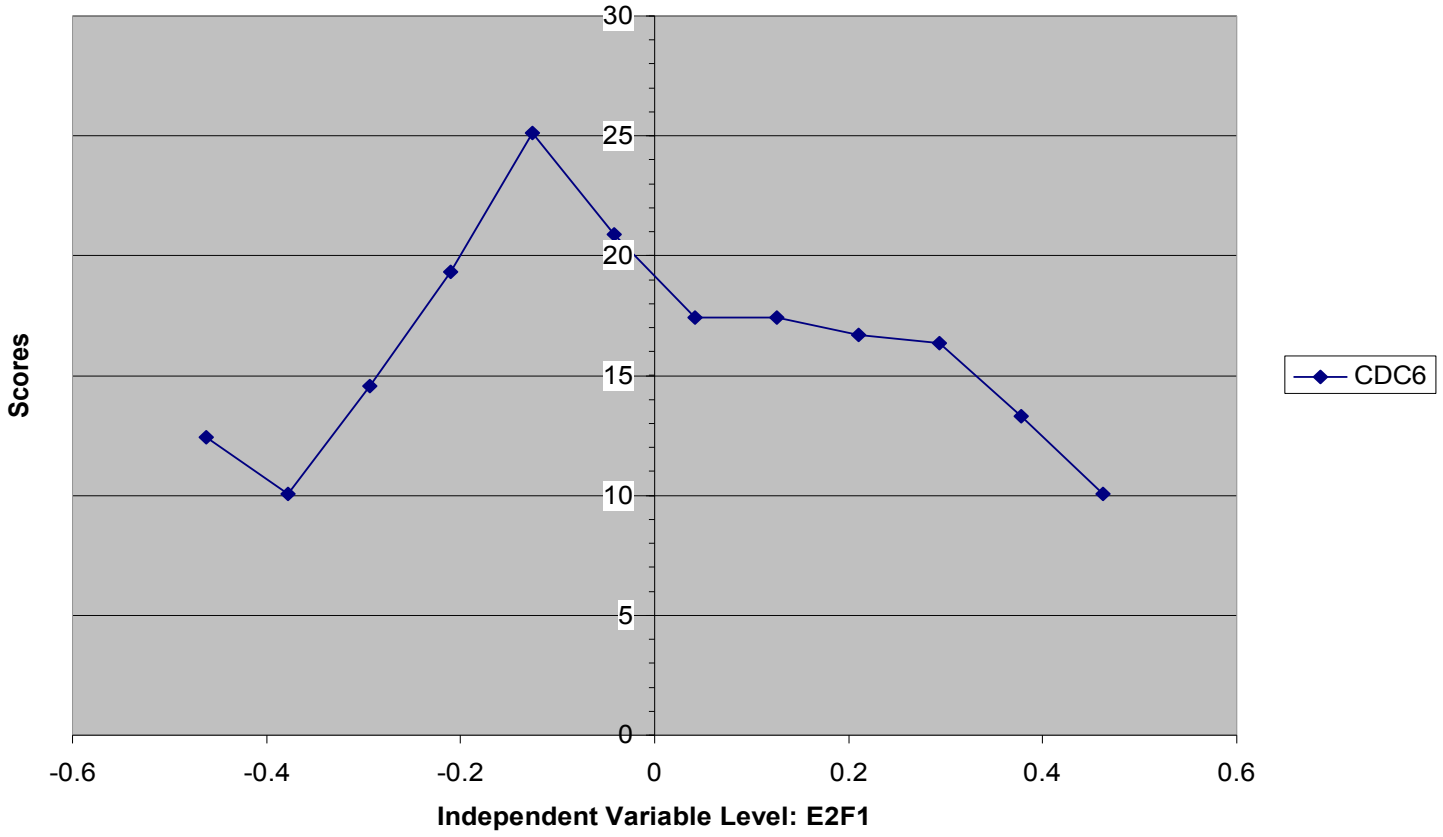
		Raw Score	P(Raw) Score	Standardized Score
19	8	-7.466667	0.00588335	-4.363133
21	0			
20	7	-3.809524	0.0494201	-2.231321
20	1			
21	6	-1.371429	0.164734	-0.810112
19	2			
22	5	-0.152381	0.284540	-0.099508
18	3			
23	4	0.152381	0.278354	0.078143
17	4			
24	3	1.371429	0.157734	0.788747
16	5			
25	2	3.809524	0.0504749	2.209955
15	6			
26	1	7.466667	0.00832004	4.341767
14	7			
27	0	12.342857	0.000539262	7.184184
13	8			

Step 7: Summarize Standardized Coordination Scores

- Select CS score with the highest absolute value to summarize
  - Each whole array
  - Any array dimension (e.g., IV level, DV level, delay, persistence, episode length, episode criterion)
  - Any combination of dimensions
- Locations of summary scores identify conditions that yield most evidence for coordination

CSs Summarized as Function of IV Level at Delay = 0

- Does DataSpeaks address non-linearity?

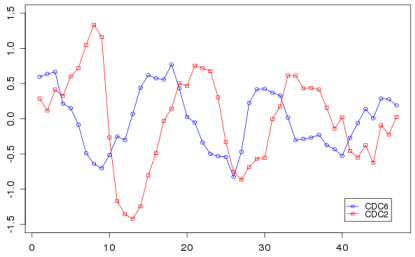


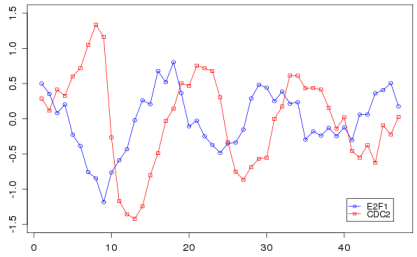
Step 8: Analyze Coordination Scores Statistically

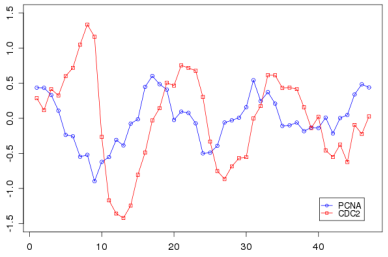
- Applies when there is more than one individual
- Describe and compare groups
- Make inferences from samples of individuals to populations
- Identify genetic and other predictors of disorder, treatment response, and differential dose requirements
- Identify CS factors to reduce dimensionality
- Inform the development of mathematical and statistical models

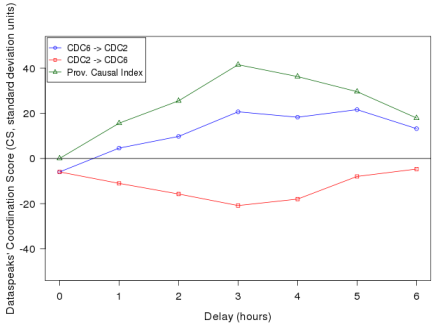
DataSpeaks Elucidates Mechanisms

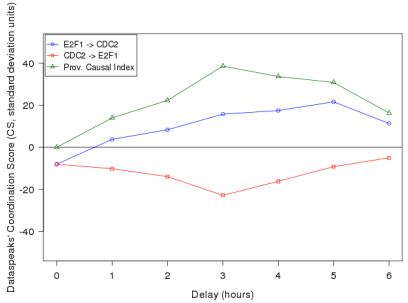
- For example, grow any of over 100 cell lines under different environmental conditions
  - Temperatures
  - Different cell culture media reagents and supplements
  - Actual or potential anti-cancer drugs in media
- Visualize detailed information about how coordination might be affected

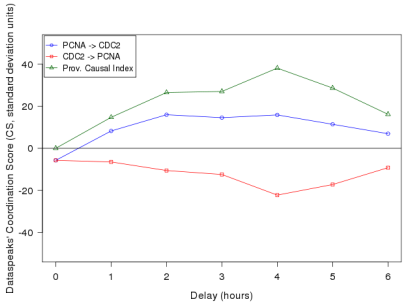

CDC6
CDC2

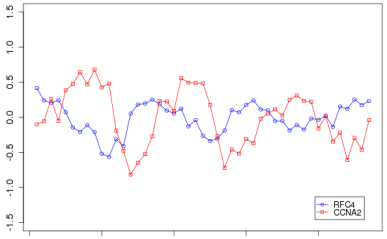

E2F1
CDC2

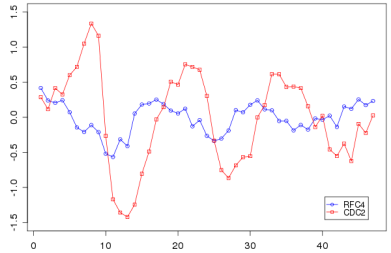

PCNA
CDC2

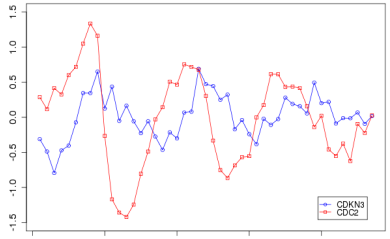

CDC6 -> CDC2
CDC2 -> CDC6
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

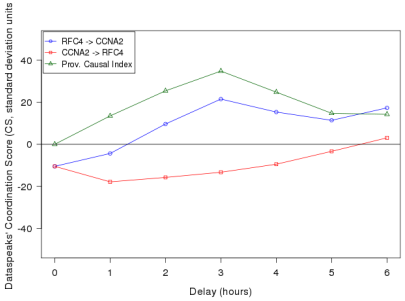

E2F1 -> CDC2
CDC2 -> E2F1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

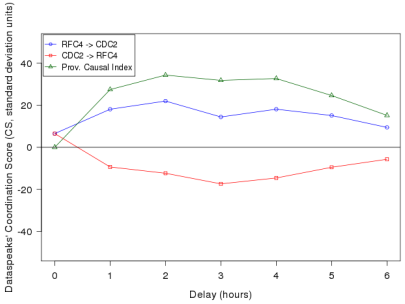

PCNA -> CDC2
CDC2 -> PCNA
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

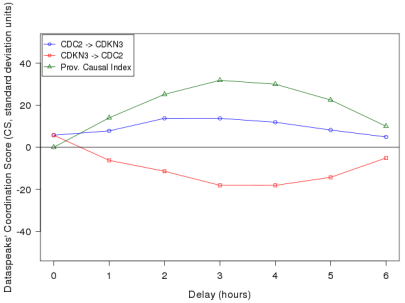

RFC4
CCNA2

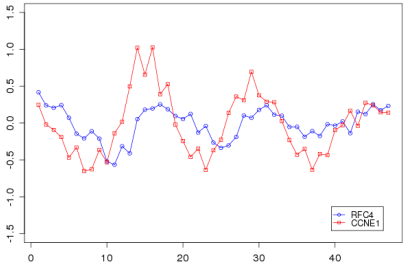

RFC4
CDC2

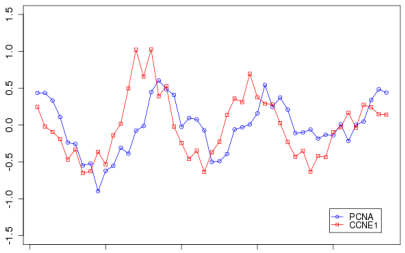

CDKN3
CDC2

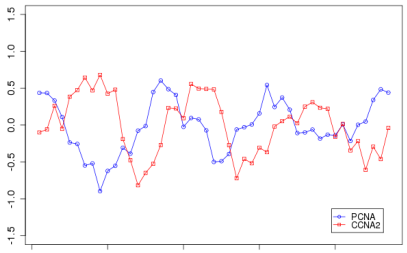

RFC4 -> CCNA2
CCNA2 -> RFC4
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

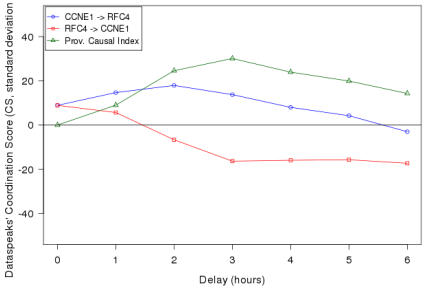

RFC4 -> CDC2
CDC2 -> RFC4
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

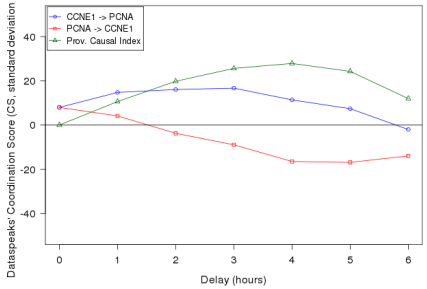

CDC2 -> CDKN3
CDKN3 -> CDC2
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

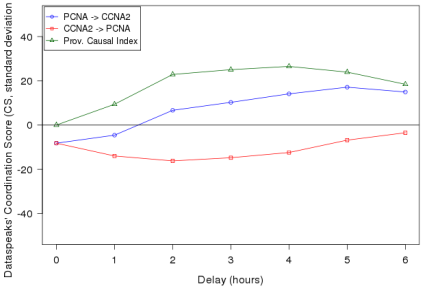

RFC4
CCNE1

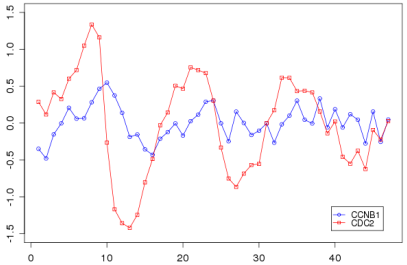

PCNA
CCNE1

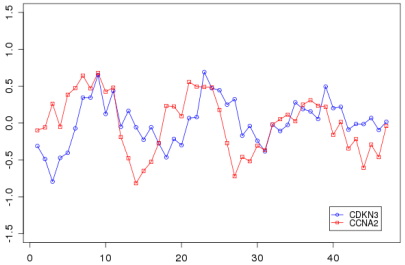

PCNA
CCNA2

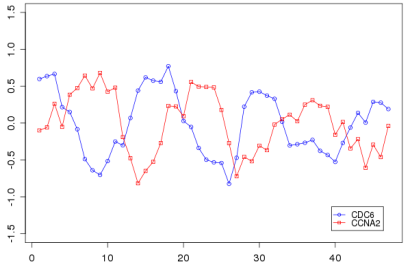

CCNE1 -> RFC4
RFC4 -> CCNE1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

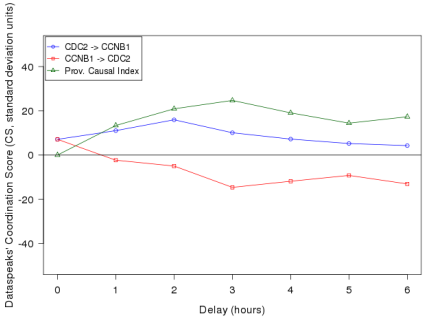

CCNE1 -> PCNA
PCNA -> CCNE1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

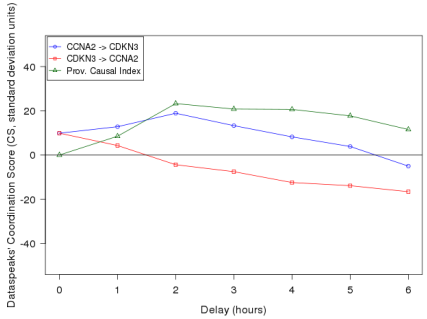

PCNA -> CCNA2
CCNA2 -> PCNA
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

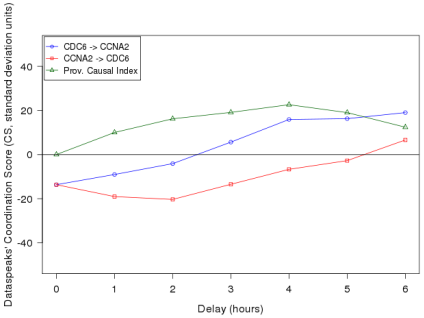

CCNB1
CDC2

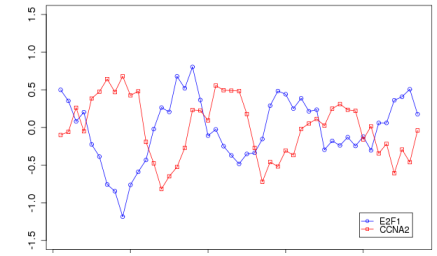

CDKN3
CCNA2

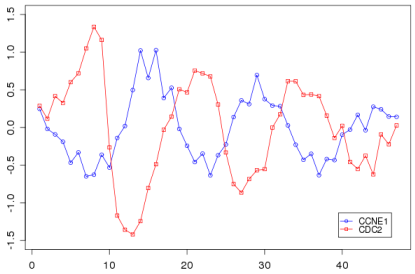

CDC6
CCNA2

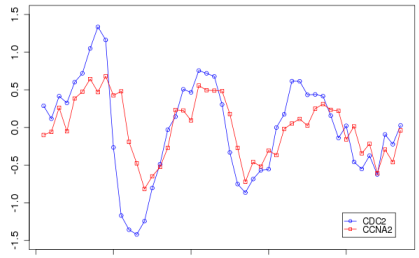

CDC2 -> CCNB1
CCNB1 -> CDC2
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

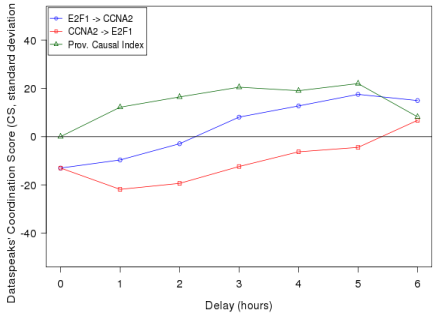

CCNA2 -> CDKN3
CDKN3 -> CCNA2
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

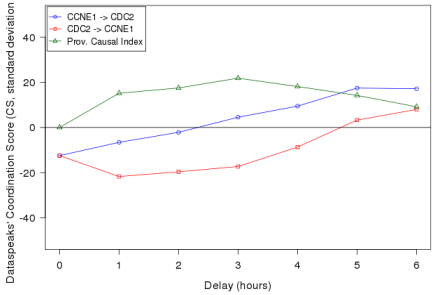

CDC6 -> CCNA2
CCNA2 -> CDC6
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

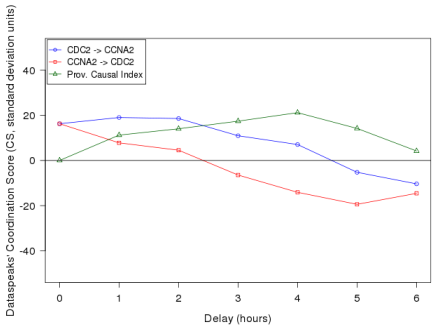

E2F1
CCNA2

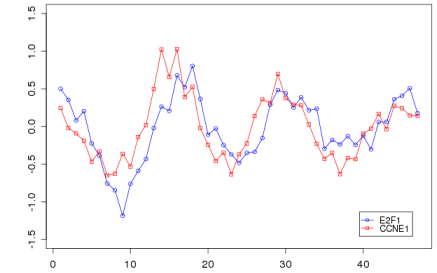

CCNE1
CDC2

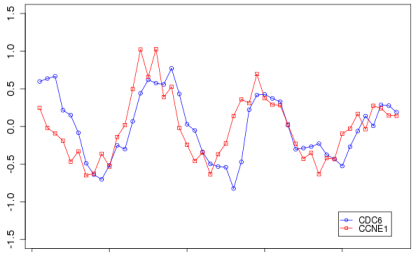

CDC2
CCNA2

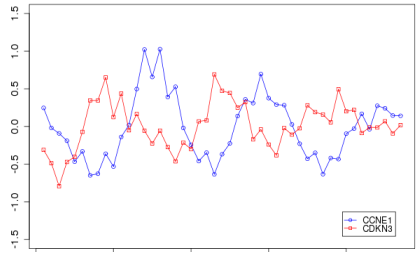

E2F1 -> CCNA2
CCNA2 -> E2F1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

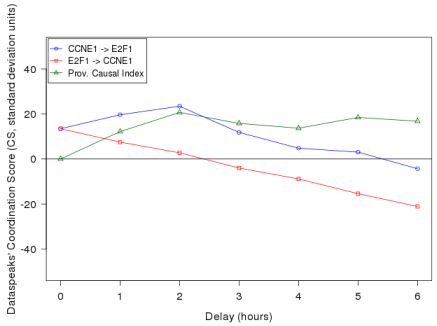

CCNE1 -> CDC2
CDC2 -> CCNE1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

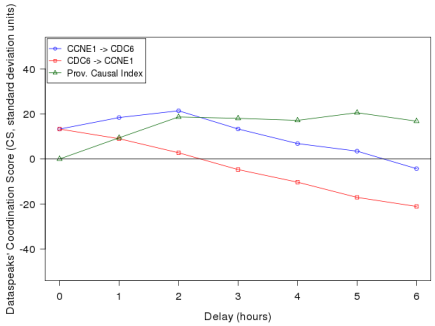

CDC2 -> CCNA2
CCNA2 -> CDC2
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

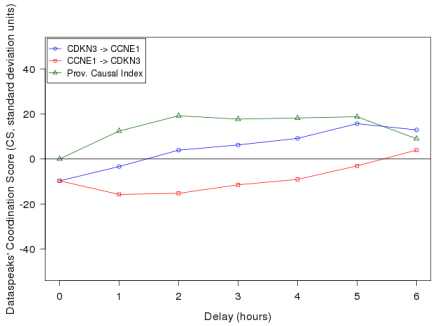

E2F1
CCNE1

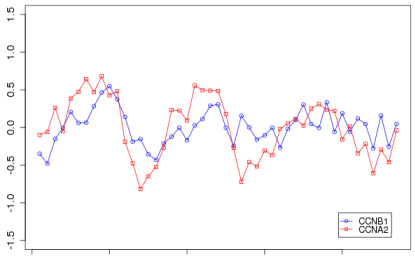

CDC6
CCNE1

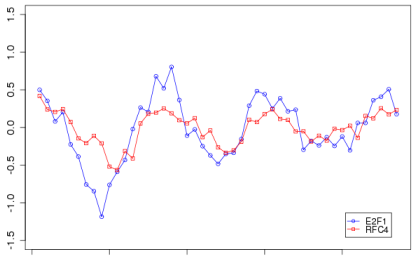

CCNE1
CDKN3

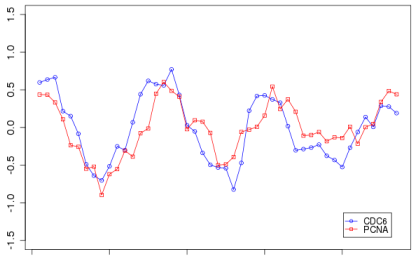

CCNE1 -> E2F1
E2F1 -> CCNE1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

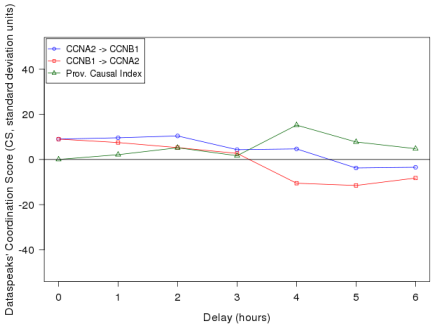

CCNE1 -> CDC6
CDC6 -> CCNE1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

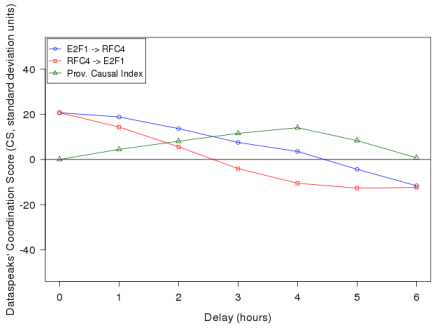

CDKN3 -> CCNE1
CCNE1 -> CDKN3
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

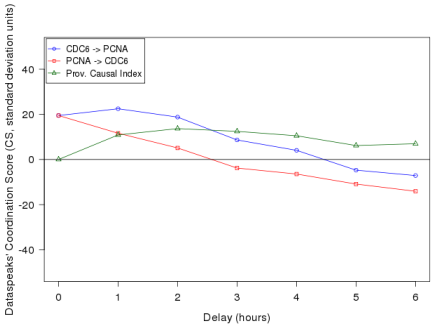

CCNB1
CCNA2

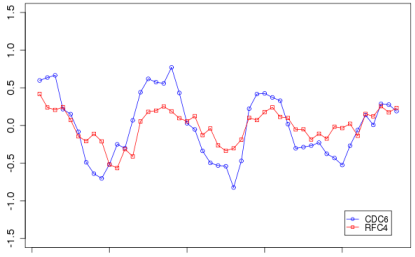

E2F1
RFC4

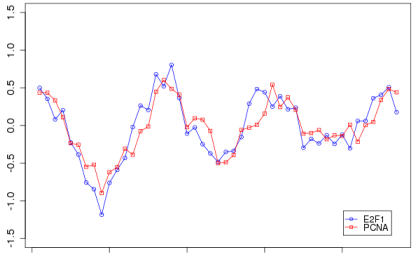

CDC6
PCNA

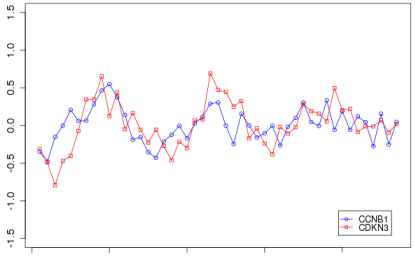

CCNA2 -> CCNB1
CCNB1 -> CCNA2
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

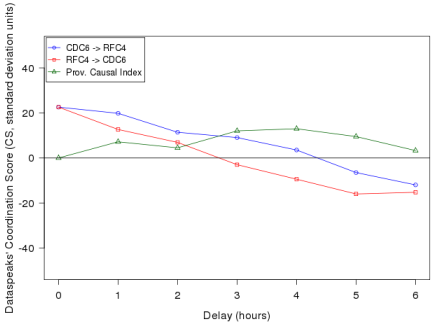

E2F1 -> RFC4
RFC4 -> E2F1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

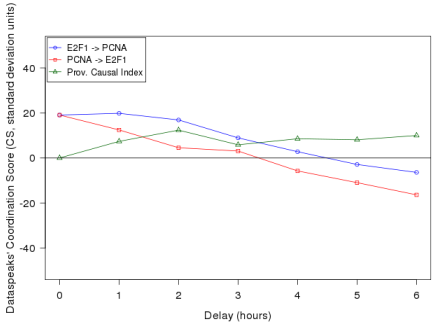

CDC6 -> PCNA
PCNA -> CDC6
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

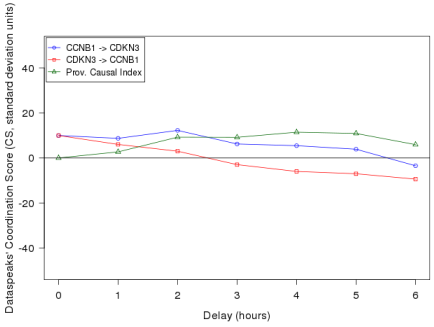

CDC6
RFC4

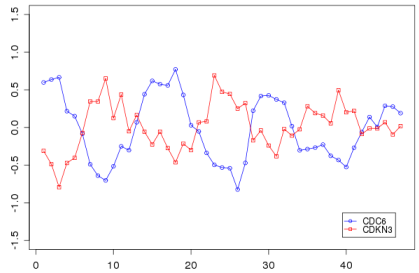

E2F1
PCNA

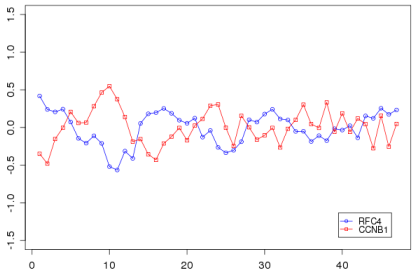

CCNB1
CDKN3

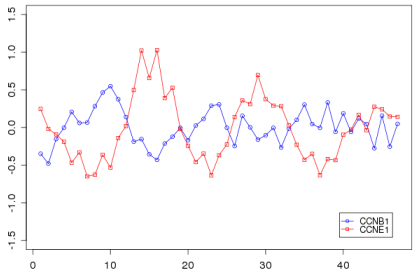

CDC6 -> RFC4
RFC4 -> CDC6
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

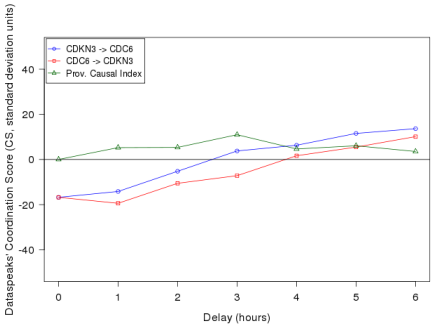

E2F1 -> PCNA
PCNA -> E2F1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

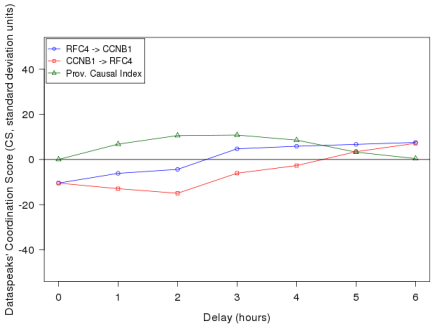

CCNB1 -> CDKN3
CDKN3 -> CCNB1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

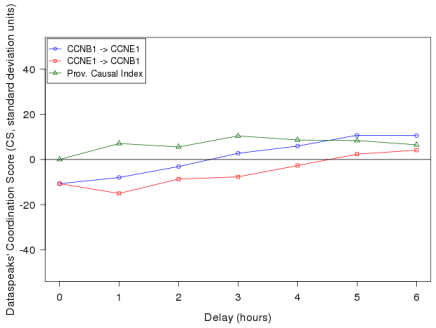

CDC6
CDKN3

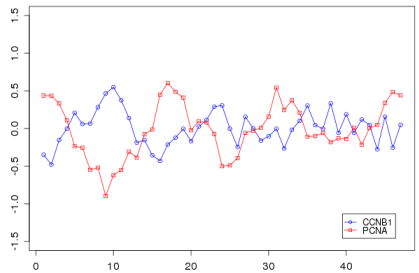

RFC4
CCNB1

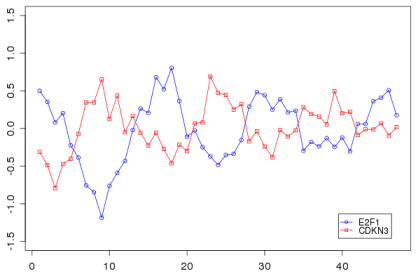

CCNB1
CCNE1

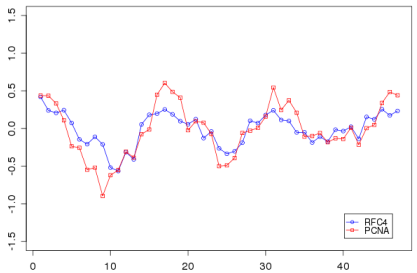

CDKN3 -> CDC6
CDC6 -> CDKN3
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

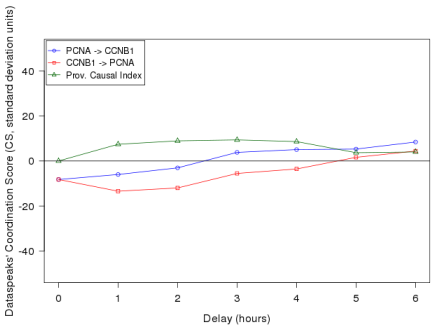

RFC4 -> CCNB1
CCNB1 -> RFC4
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

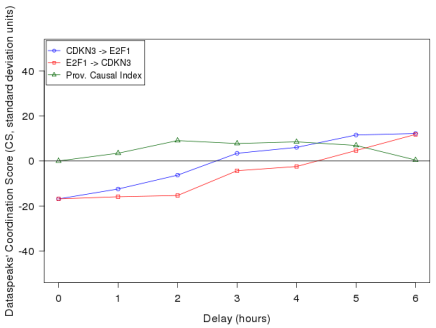

CCNB1 -> CCNE1
CCNE1 -> CCNB1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

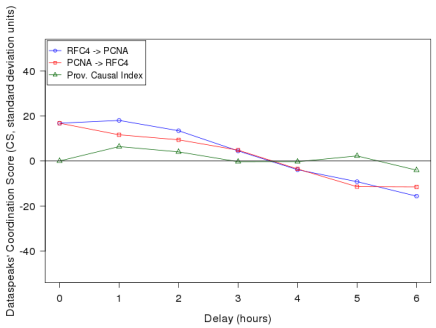

CCNB1
PCNA

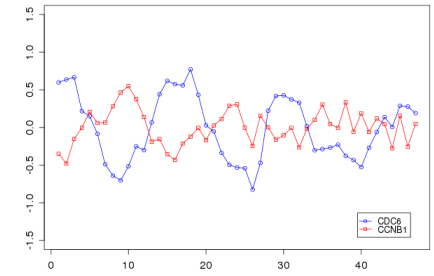

E2F1
CDKN3

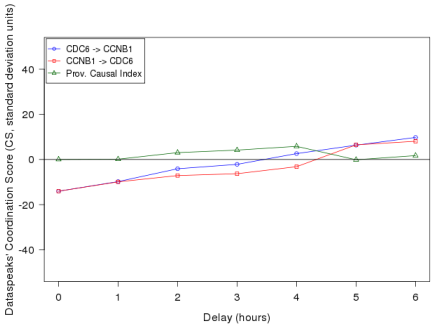

RFC4
PCNA

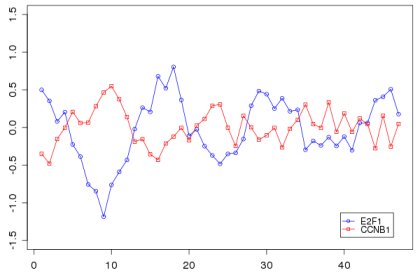

PCNA -> CCNB1
CCNB1 -> PCNA
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

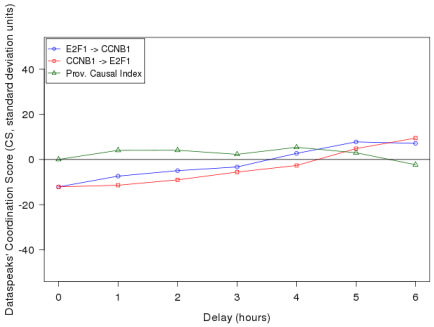

CDKN3 -> E2F1
E2F1 -> CDKN3
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

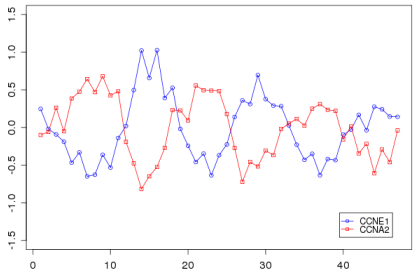

RFC4 -> PCNA
PCNA -> RFC4
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

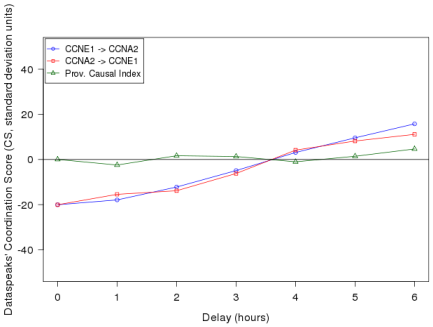

CDC6
CCNB1

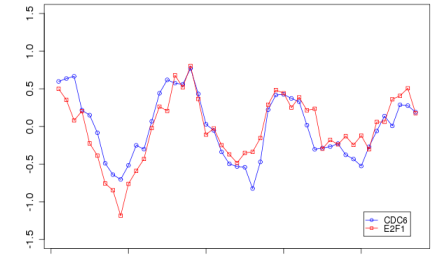

CDC6 -> CCNB1
CCNB1 -> CDC6
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

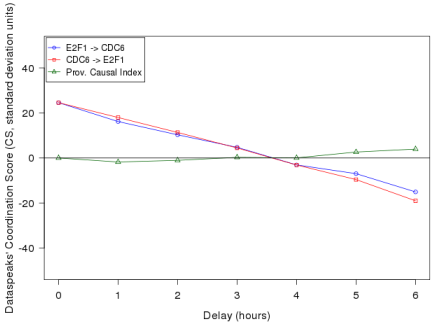

E2F1
CCNB1

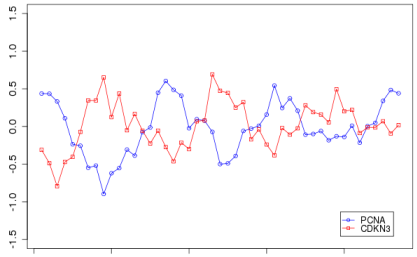

E2F1 -> CCNB1
CCNB1 -> E2F1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

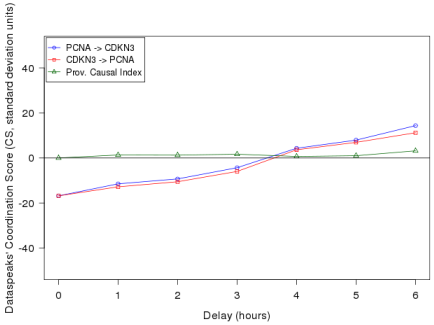

CCNE1
CCNA2

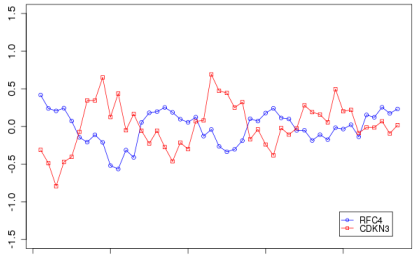

CCNE1 -> CCNA2
CCNA2 -> CCNE1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

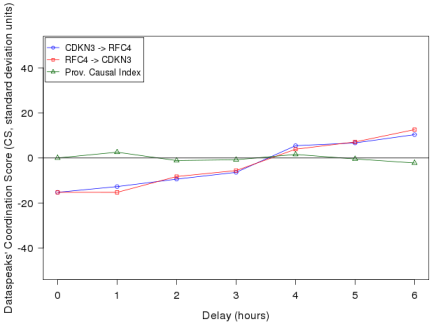

CDC6
E2F1


E2F1 -> CDC6
CDC6 -> E2F1
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)


PCNA
CDKN3


PCNA -> CDKN3
CDKN3 -> PCNA
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)


RFC4
CDKN3


CDKN3 -> RFC4
RFC4 -> CDKN3
Prov. Causal Index
Dataspeaks' Coordination Score (CS, standard deviation units)
Delay (hours)

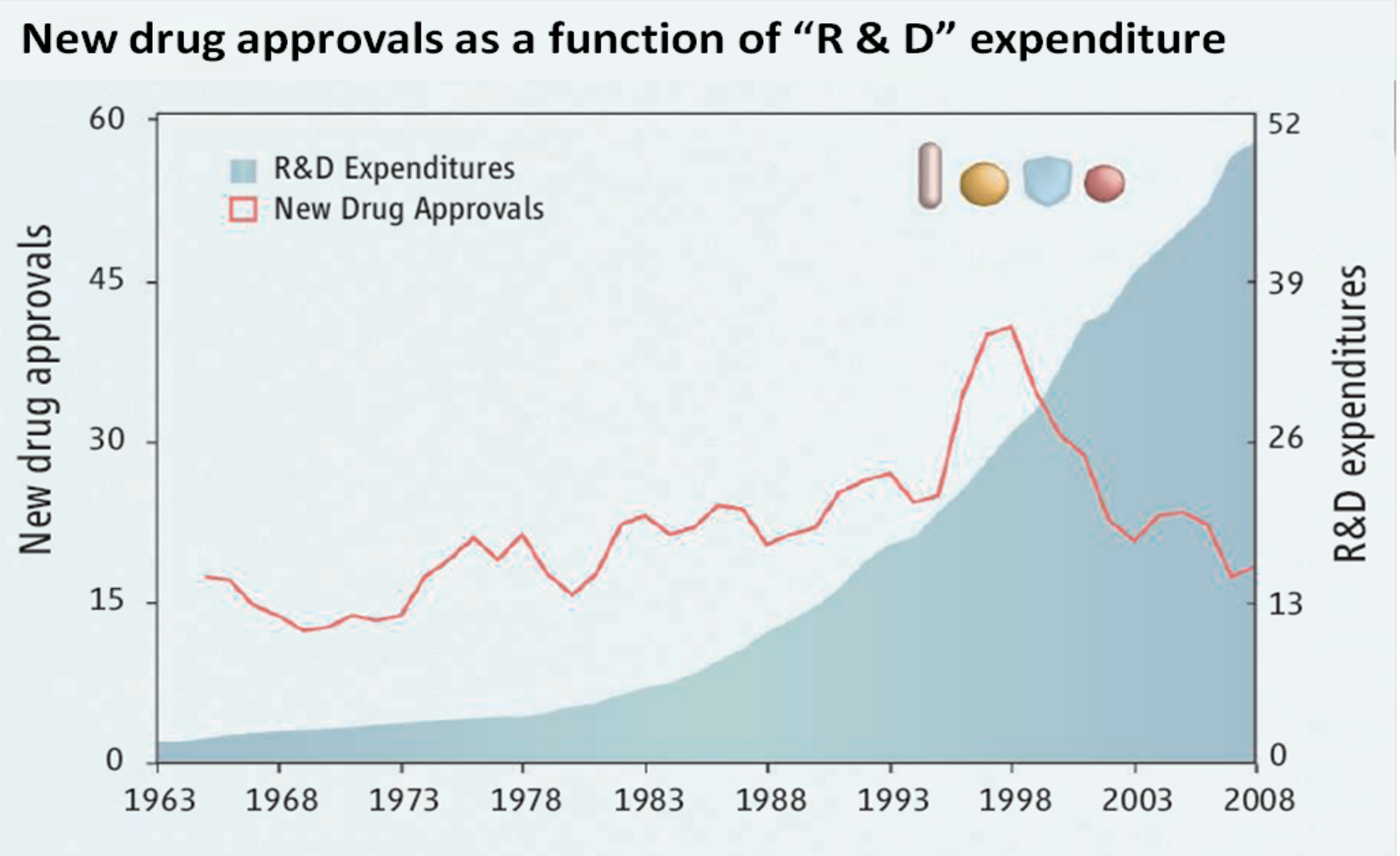
DataSpeaks Will Advance Medical Diagnosis

- Many chronic health problems appear to be *disorders of coordinated action* at various levels
  - Biological (proteins, lipids, carbohydrates, metabolites, electrophysiological, etc.)
  - Psychological (mental and physical behavior)
  - Social (e.g., family and work role performance)
- Coordinated action is an emergent system property
- Dream project: Apply DataSpeaks to BOLD fMRI data to visualize disorders of functional connectivity between and among brain regions

Let's Rock the RCT Design Boat

- Combine DataSpeaks with randomized experimental control *exercised over time* for *individuals*
- Tactical win now for drug rescue and repurposing
- Potential to become new gold standard for many RCTs when drugs are developed and used to manage or control chronic health disorders
- Use Ultra RCT designs and DataSpeaks' Software as a Service (SaaS) to measure benefit and harm over time and across response variables for each individual
- Estimate 10% of the cost and 50% faster and 1000 times safer !!!

“Houston, we have a problem”



Houston, we have another problem

- Drug safety problem
  - About 100,000 deaths and 2,000,000 hospitalizations per year in U.S.
  - Vioxx, Bextra – product withdrawals
  - Multi billion dollar legal liability due largely to safety problems that derive from weak science that leaves room for error and bad behavior
  - Pfizer➔NCRC

Diagnose *Scientific* Causes of RCT Design Problem

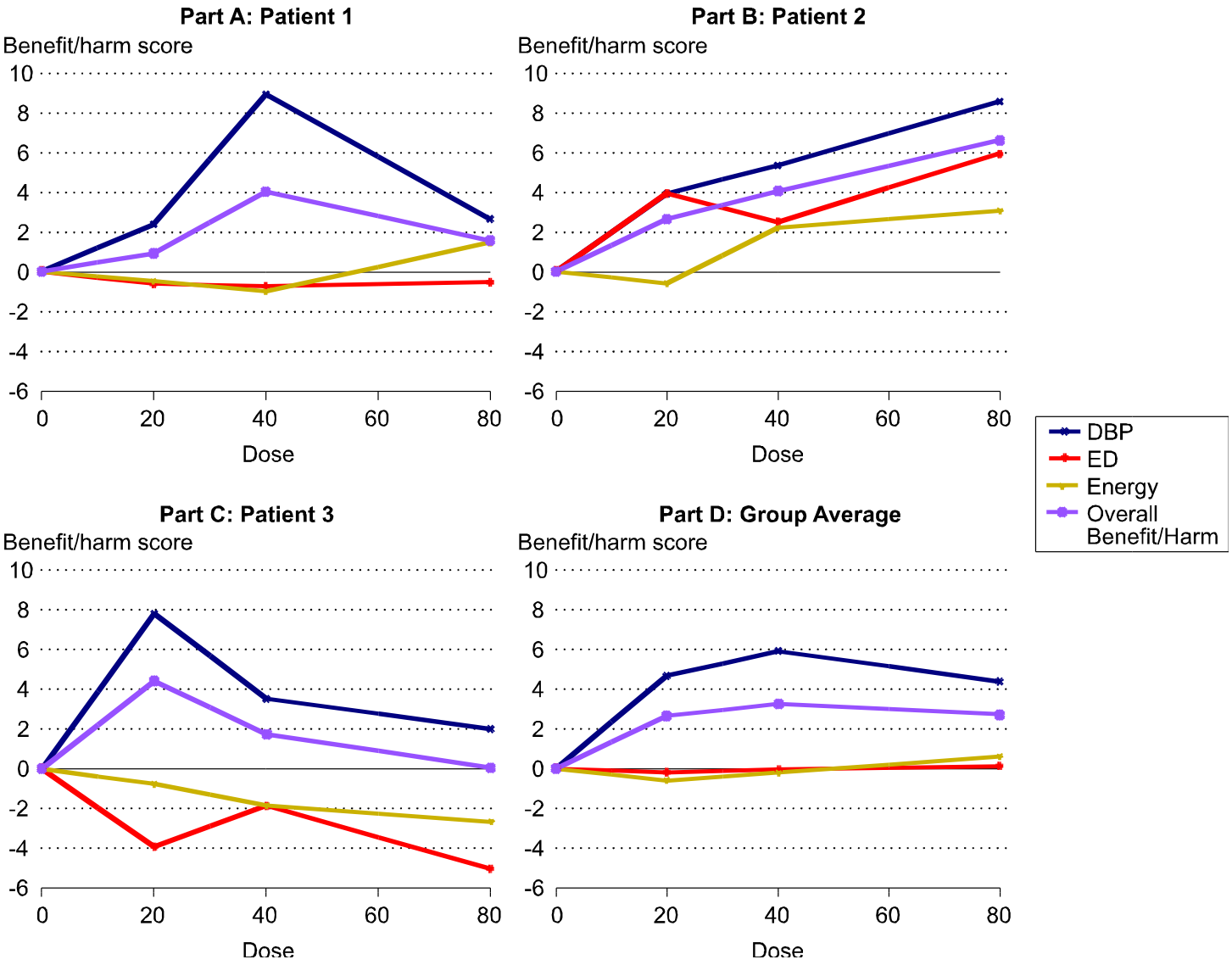
- Current first-generation RCT designs, which are for *causality assessment,* date back to the 1940s
  - Streptomycin
  - ENIAC
- Four types of confounding
  - Individuality with measurement error
  - True responders with placebo responders
  - Dose with type of treatment
  - Treatment effects with how they are valued
- Too much data is going to waste

Elephants in the RCT Design Room

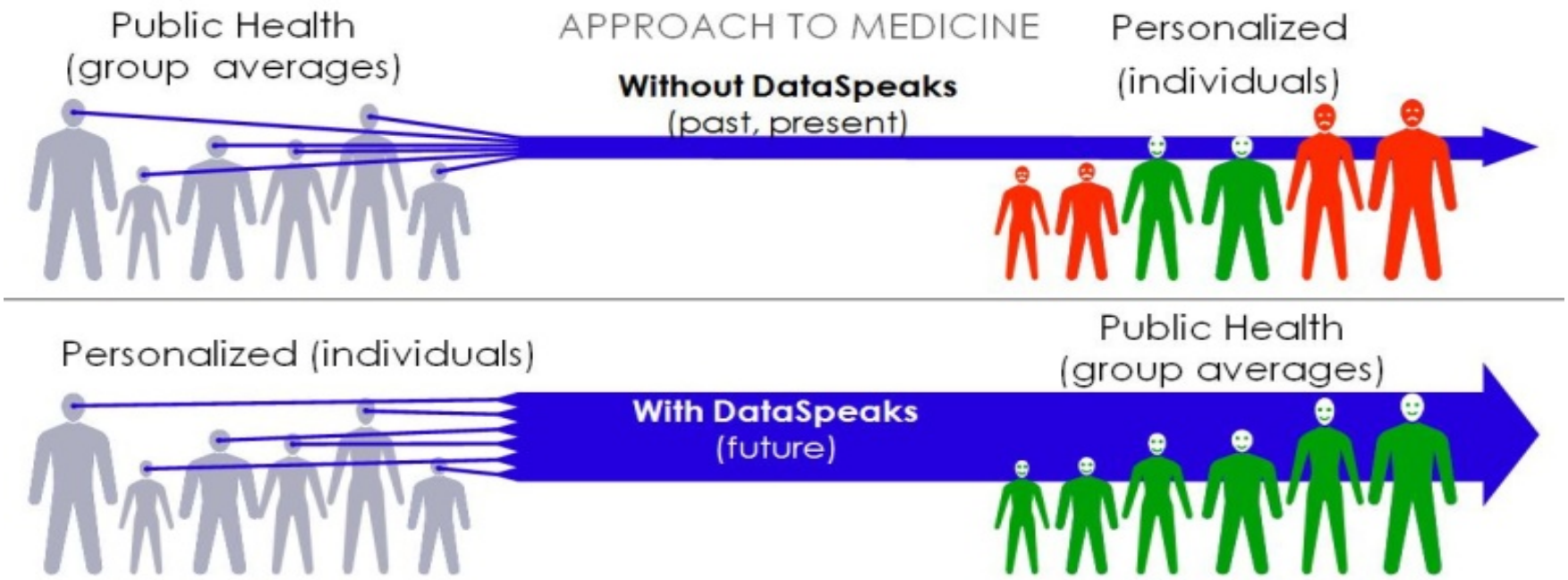
Ultra RCT Example, Mock Data

Patient Variable	Week 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Interaction score	Direction[1]	Benefit/ Harm Score	Weight[2]	Overall Benefit/ Harm Score
	Pair 1				Pair 2				Pair 3				Pair 4								
	Period				Period				Period				Period								
	1		2		1		2		1		2		1		2						
Patient 1																					
Dose	20	20	40	40	0	0	40	40	40	40	20	20	80	80	40	40					
DBP[3]	96	98	85	81	91	96	84	87	80	78	93	98	82	77	81	78	-8.92	-	**8.92**	4	
ED[4]	3	2	2	3	2	1	2	1	3	2	1	2	3	2	2	2	0.74	-	**-0.74**	2	
Energy	4	3	4	5	4	3	4	3	2	4	3	3	4	4	2	3	1.46	+	**1.46**	2	**4.64**
Patient 2																					
Dose	0	0	20	20	40	40	0	0	20	20	80	80	80	80	40	40					
DBP	98	91	89	88	89	84	96	98	89	93	86	80	76	75	80	92	-8.56	-	**8.56**	4	
ED[4]	2	1	2	2	3	2	1	2	2	3	2	4	4	3	4	2	5.93	+	**5.93**	1	
Energy	2	3	2	2	3	1	2	3	2	2	3	3	4	3	2	3	3.05	+	**3.05**	2	**6.61**
Patient 3																					
Dose	40	40	20	20	0	0	80	80	20	20	40	40	20	20	80	80					
DBP	76	79	74	80	88	90	78	68	75	77	81	78	73	79	76	82	-7.81	-	**7.81**	4	
ED[4]	1	3	2	2	1	1	2	4	2	2	2	1	3	2	3	5	5.04	-	**-5.04**	1	
Energy	4	3	2	2	3	4	3	2	5	3	4	3	4	4	1	2	-2.67	+	**-2.67**	1	**3.92**
Group Average																					**5.06 t=6.29 p=.0242**

Get Doses Right



Harmonize P4 Medicine with Public Health Approach to Medicine



Personalized medicine that improves individual health will improve group average or public health with the help of statistics.

What problem could not be done better with multiple time series?

Forensic identification with DNA.

Most problems involving living systems, which are CDAS that manifest emergent system properties, can be done better with time series and DataSpeaks' software.

Next Steps, Collaborate?

- Publications
  - http://dataspeaks.com/resources/APA-JCCP-1992-Vol60-No2-P225-239.pdf
  - Patents
  - Need new, co-authored, peer-reviewed publication with a version of the algorithm that is
    - More mathematically elegant
    - Computationally efficient?
    - Illustrated in a particular scientific context
- Faculty and Graduate Student Projects

Next Steps, Collaborate?

- Software development
- In silico simulations as with Ultra RCTs
- J&J, tranSMART?
- Grants, e.g., SBIR
  - Phenotyping Pain Treatment Responses (NIH PhenX Project)
  - Drug Rescue by Data Rescue from RCTs
  - Diagnosing Functional Brain Disorders with a New Algorithm for Processing Functional Brain Imaging Data
- Commercialization

Thank you

- Please contact Curt Bagne
  - cbagne@DataSpeaks.com
  - 248 952-1968 (home phone)
  - 2971 Vineyards Drive, Troy, MI 48098