V2: Circadian rhythms, time-series analysis (intro)

A circadian gene expression atlas in mammals: Implications for biology and medicine

Ray Zhang[a,1], Nicholas F. Lahens[a,1], Heather I. Ballance[a], Michael E. Hughes[b,2], and John B. Hogenesch[a,2]

[a]Department of Pharmacology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; and [b]Department of Biology, University of Missouri, St. Louis, MO 63121

Introduction: 3 paragraphs

(1) What are circadian rhythms? Biological/medical relevance

(2) Previous work, only single organs analyzed – here: profiling of 12 organs.

(3) What has been achieved in this study?

Methods section:

(1) Animal Preparation and Organ Collection

(2) Microarray Data

(3) RNA-seq Data

(4) Oscillation Detection

question: why microarray and RNA-seq?

Oscillation detection: JTK_CYCLE

JTK_CYCLE applies the Jonckheere-Terpstra-Kendall (JTK) algorithm to alternative hypothesized group orderings corresponding to a range of user-defined period lengths and phases.

JTK is a special case of Kendall's more general method of rank correlation.

In effect, the JTK_CYCLE algorithm finds the optimal combination of period and phase that minimizes the exact p-value of Kendall's tau correlation between an experimental time series and each tested cyclical ordering.

For the ease of interpretation, group orderings are derived from **cosine curves**, although generally speaking, the choice of group order can be anything.

Each minimal p-value is Bonferroni-adjusted for multiple testing.

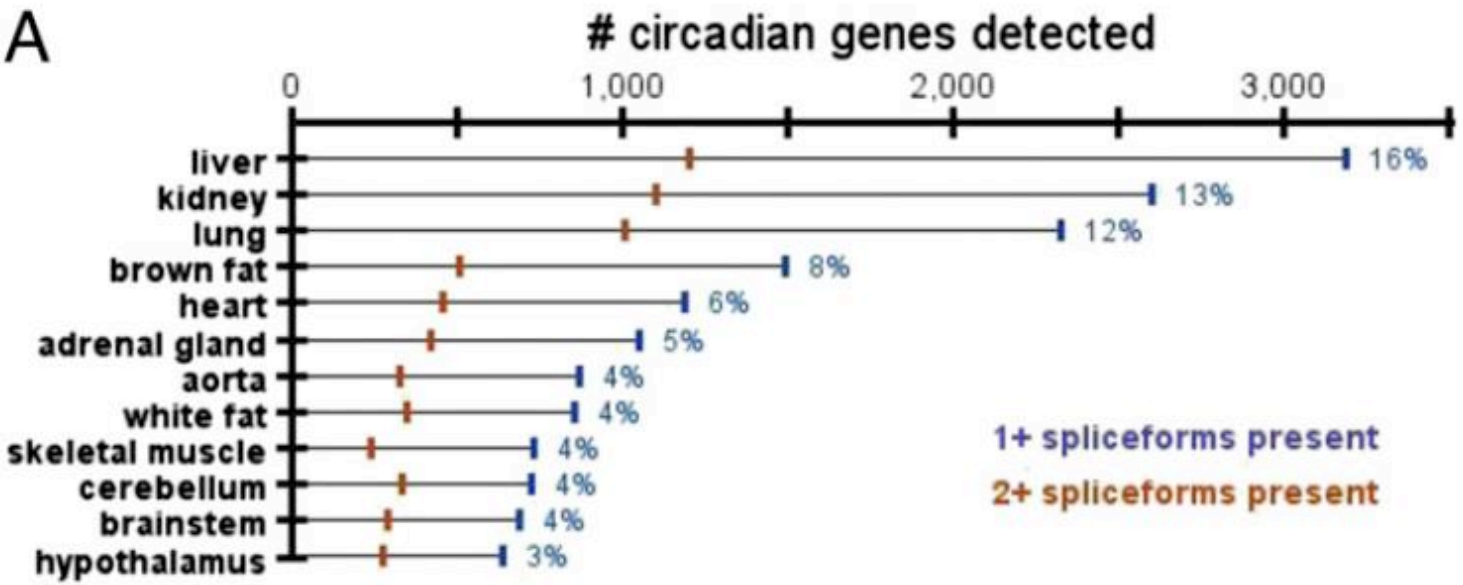
J Biol Rhythms. 2010;25:372-80.

Results: start with overview of the data …
How many circadian genes are detected in various organs?

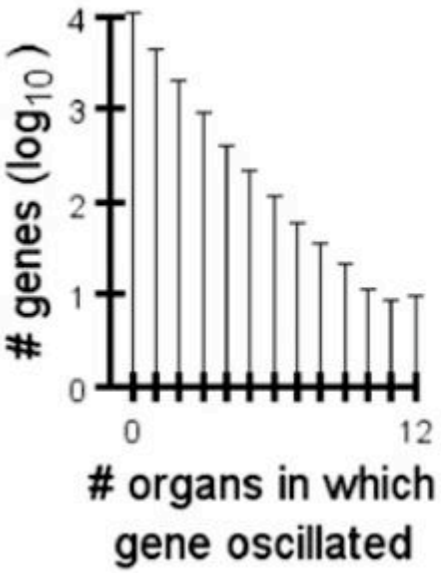
A circadian gene expression atlas in mammals: Implications for biology and medicine

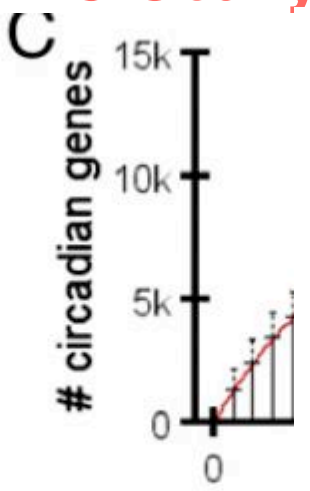
Ray Zhang[a,1], Nicholas F. Lahens[a,1], Heather I. Ballance[a], Michael E. Hughes[b,2], and John B. Hogenesch[a,2]

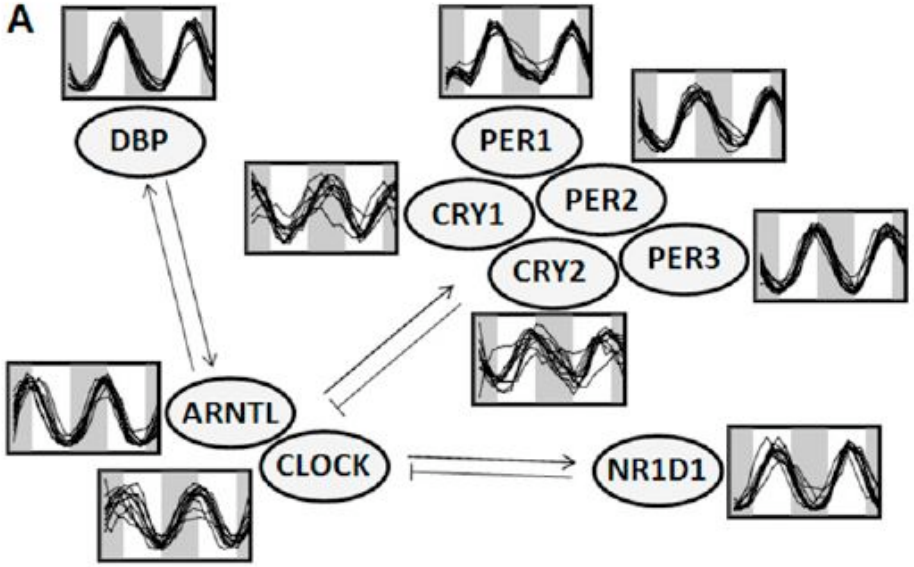
[a]Department of Pharmacology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; and [b]Department of Biology, University of Missouri, St. Louis, MO 63121



Globally oscillating genes







Only 10 genes oscillated in all organs:

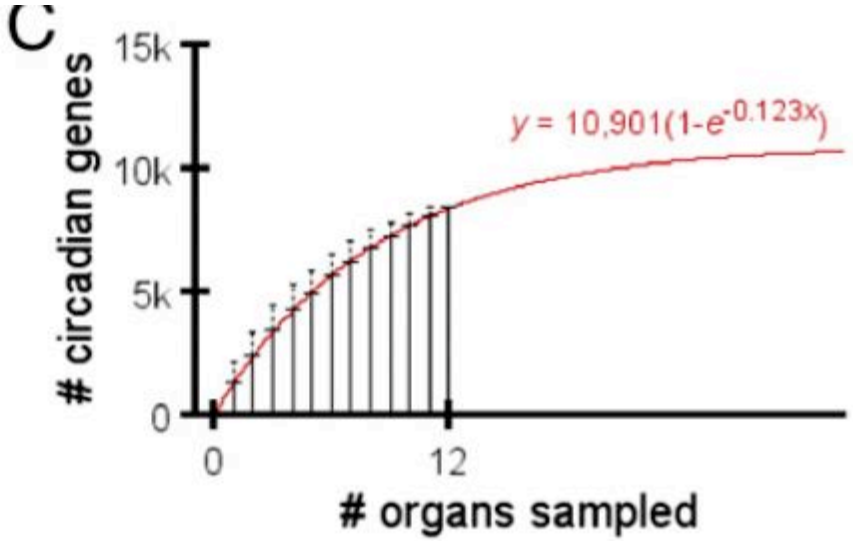
Arntl, Dbp, Nr1d1, Nr1d2, Per1, Per2, and *Per3* (core clock factors – **as expected**), and *Usp2*, *Tsc22d3*, and *Tspan4*.

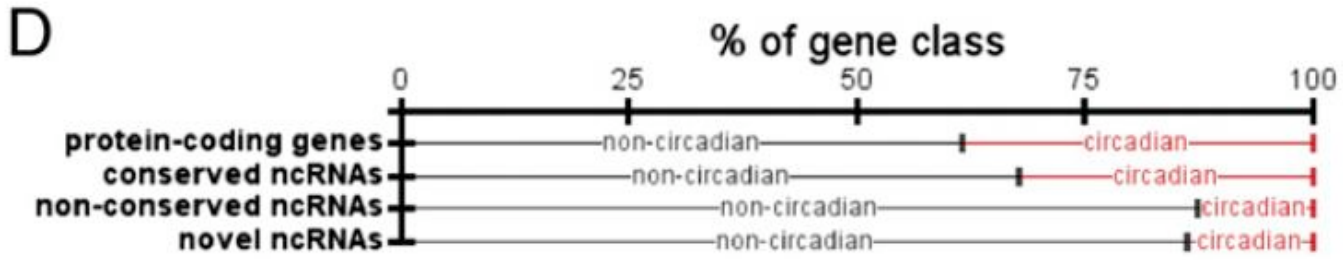
Usp2 - Ubiquitin carboxyl-terminal hydrolase 2

Tsc22d3 - TSC22 domain family protein 3

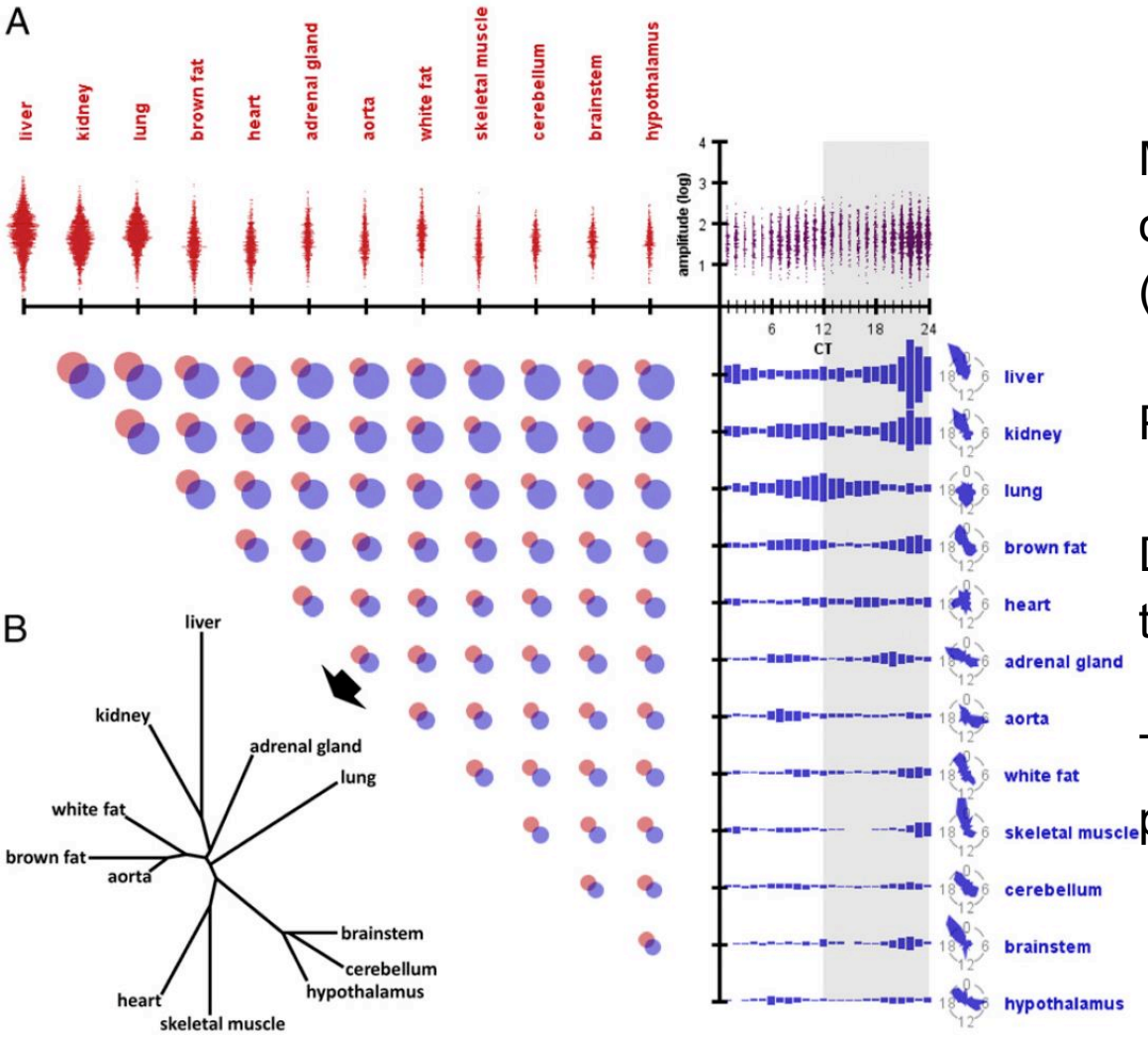
Tspan4 - The protein encoded by this gene is a member of the transmembrane 4 superfamily, also known as the tetraspanin family.

Overlap of genes/organs (B), how many expected (C)?





(A) Phases + overlap, (B) similarity



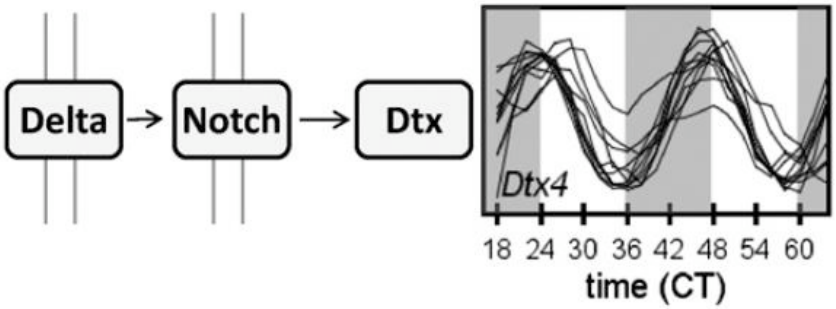
Most circadian genes show organ-specific expression (small overlap).

Peaks often at dawn and dusk.

Developmentally related organs tend to share circadian genes.

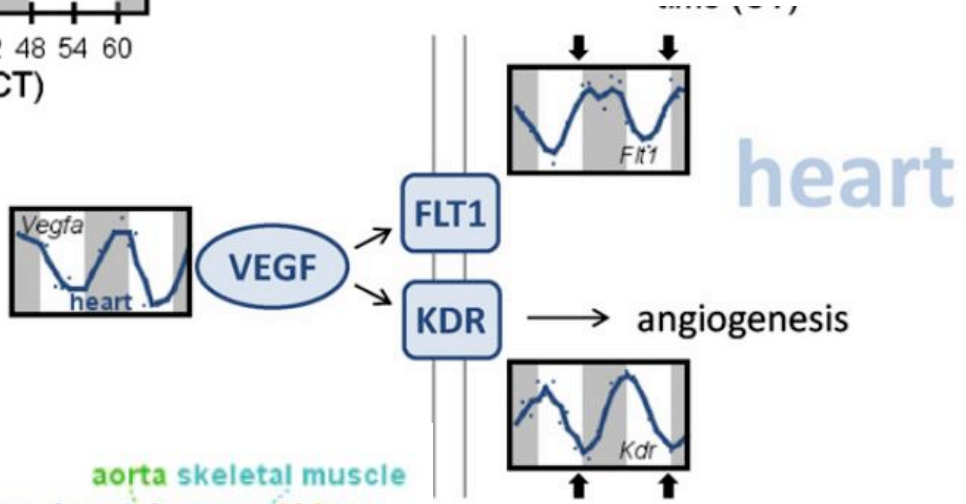
Tree generated by similarity of peak phases.

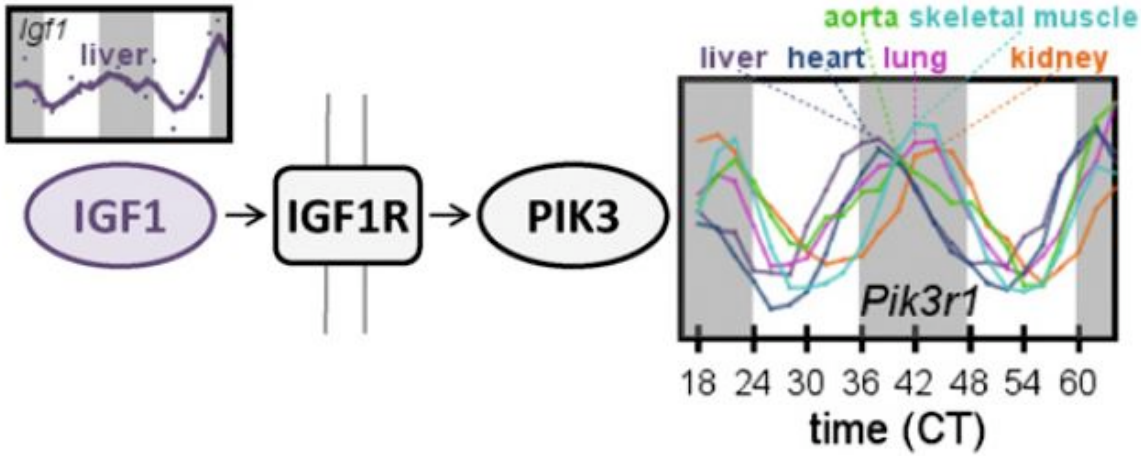
Examples



Dtx4, a Notch pathway E3 ubiquitin ligase, oscillated in phase with *Arntl* in all organs

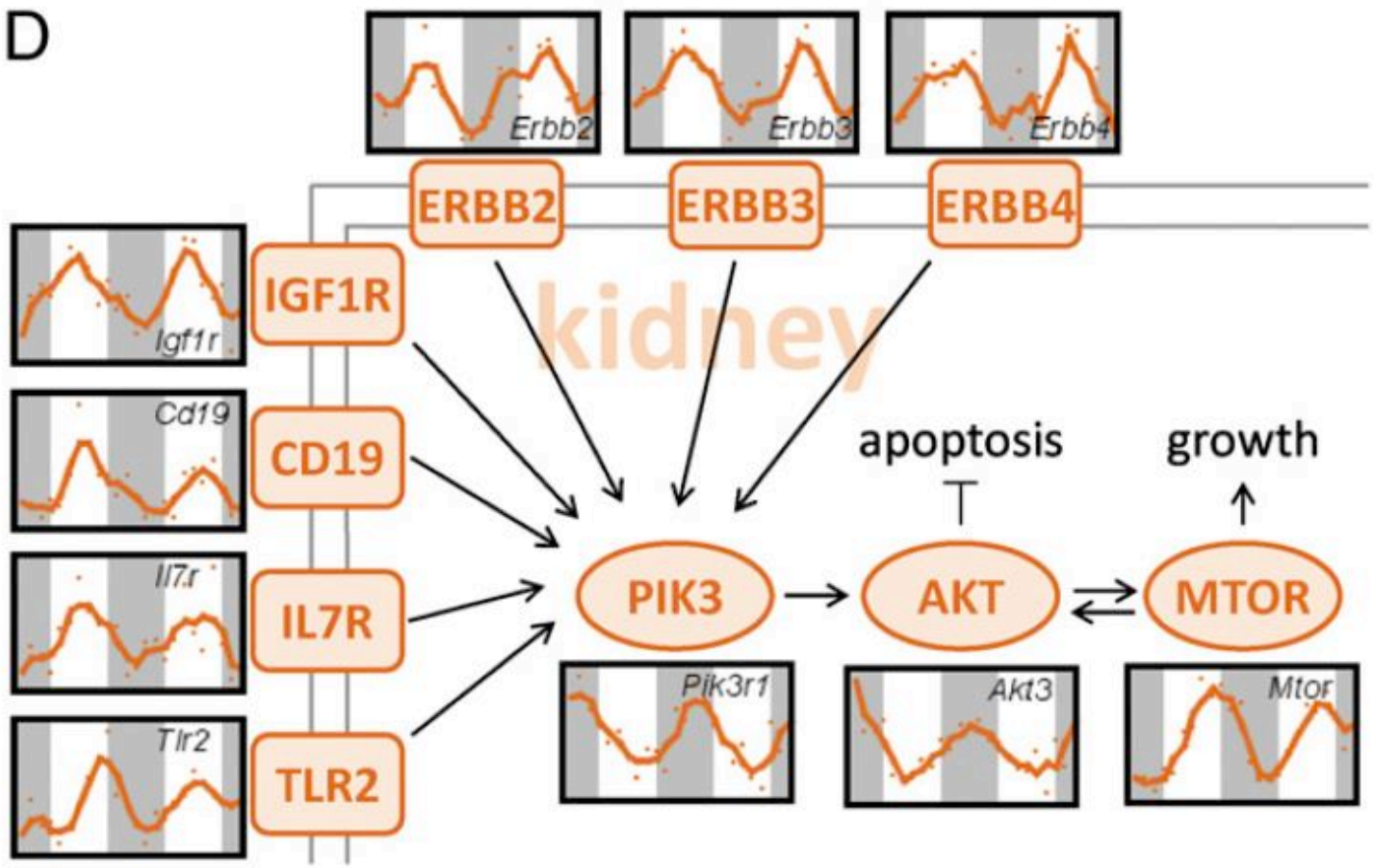
Two VEGF-receptors are expressed alternatively.





IGF1 is most produced in liver -> peaks at the same time throughout body, however PIK3r1 receptor peaks at different times in organs.

Multiple coordinated pathways control PIK3-AKT-MTOR



Multiple synchronous receptors feed into PIK3-AKT-MTOR pathway that controls growth and apoptosis

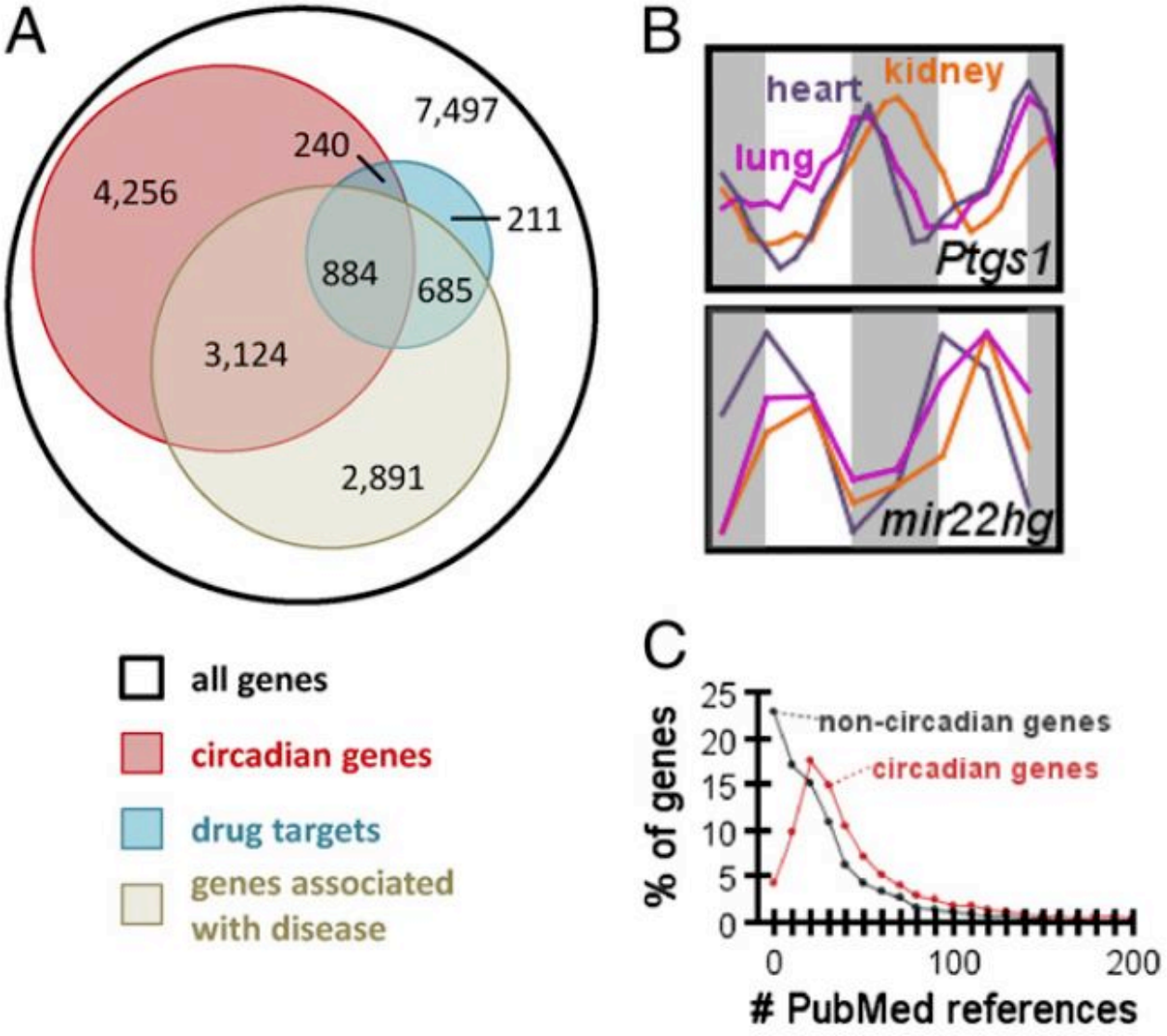
Relevance: mouse -> humans, drugs

Table 1. Drugs of the top-100 best-seller list that target circadian genes and have half-life < 6h

Rank	Sales, $	Trade name	Indications	Circadian-gene targets	Organs in which targets oscillate
2	1.46 b	Nexium	Gastritis, GERD, Esophagitis	*Atp4a*	L
5	1.28 b	Advair Diskus	Asthma, Chronic obstructive pulmonary di...	*Serpina6, Pgr, Nr3c2, Adrb2, Pla2g4a*	Lu, H, L, K, S, A
11	794 m	Rituxan	Rheumatoid arthritis, Non-Hodgkin's lymp...	*Fcgr2b, Ms4a1, Fcgr3*	L, K, S
20	538 m	Diovan	Hypertension, Heart failure	*Slc22a6, Agtr1a, Slco1b2, Car4, Kcnma...*	H, AG, L, K, S
27	431 m	Vyvanse	Attention deficit hyperactivity disorder	*Adra1b*	L
32	392 m	Tamiflu	Influenza	*Neu2, Neu1, Ces1g, Slc22a8, Slc15a1, ...*	Lu, L, BF, K, C
33	383 m	Ritalin	Attention deficit hyperactivity disorder	*Slc6a4*	AG, K
37	348 m	AndroGel	Hypogonadism	*Slc22a4, Slc22a3, Ar, Cyp1a1, Cyp2b10...*	Lu, H, BS, WF, AG...
38	346 m	Lidoderm	Pain	*Slc22a5, Cyp2b10, Egfr, Abcb1a*	Lu, H, AG, BF, L,...
44	304 m	Seroquel XR	Bipolar disorder, Major depressive disor...	*Htr2c, Htr1b, Htr2a, Chrm2, Drd4, Adr...*	Lu, H, BS, WF, AG...
45	289 m	Viagra	Erectile dysfunction	*Cyp1a1, Pde6g, Abcc5, Abcc10, Pde5a, ...*	Lu, H, BS, WF, AG...
47	281 m	Niaspan	Hyperlipidemia	*Slco2b1, Slc22a5, Qprt, Slc16a1*	Lu, H, BS, AG, WF...
48	279 m	Humalog	Diabetes mellitus T2	*Igf1r*	K
49	274 m	Alimta	Mesothelioma, Nonsmall cell lung cancer	*Tyms, Atic, Gart, Slc29a1*	Lu, H, BS, BF, L,...
54	267 m	Combivent	Asthma, Chronic obstructive pulmonary di...	*Slc22a5, Slc22a4, Chrm2, Adrb1, Adrb2*	Lu, H, BS, BF, K,...
56	262 m	ProAir HFA	Asthma, Chronic obstructive pulmonary di...	*Adrb1, Adrb2*	Lu, K, S
62	240 m	Janumet	Diabetes mellitus T2	*Slc47a1, Slc22a2, Prkab1, Abcb1a, Dpp4*	H, BS, AG, Hy, L,...
66	236 m	Toprol XL	Hypertension, Heart failure	*Slc22a2, Adrb1, Adrb2, Abcb1a*	Lu, H, AG, BF, L,...
71	220 m	Vytorin	Hyperlipidemia	*Hmgcr, Cyp2b10, Soat1, Abcc2, Anpep, ...*	Lu, H, BS, AG, BF...
78	209 m	Aciphex	Gastritis, GERD, Esophagitis	*Cyp1a1, Atp4a, Abcg2*	Lu, H, BS, WF, L,...
90	189 m	Lunesta	Insomnia	*Ptgs1, Tspo, Gabra3*	Lu, H, AG, K
98	173 m	Prilosec	Gastritis, GERD, Esophagitis	*Cyp1a1, Atp4a, Abcg2, Cyp1b1, Abcb1a*	Lu, H, BS, WF, AG...
99	171 m	Focalin XR	Attention deficit hyperactivity disorder	*Slc6a4*	AG, K

Rank and sales are based on USA 2013 Q1 data from Drugs.com. A, aorta; AG, adrenal gland; BF, brown fat; BS, brainstem; C, cerebellum; H, heart; Hy, hypothalamus; K, kidney; L, liver; Lu, lung; S, skeletal muscle; WF, white fat.

How many are drug-target related?

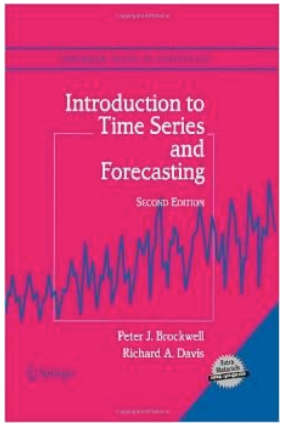


How to proceed?

Speculate what these authors will do next …

Introduction to time series analysis

After book of Peter Brockwell &
Richard A. Davis



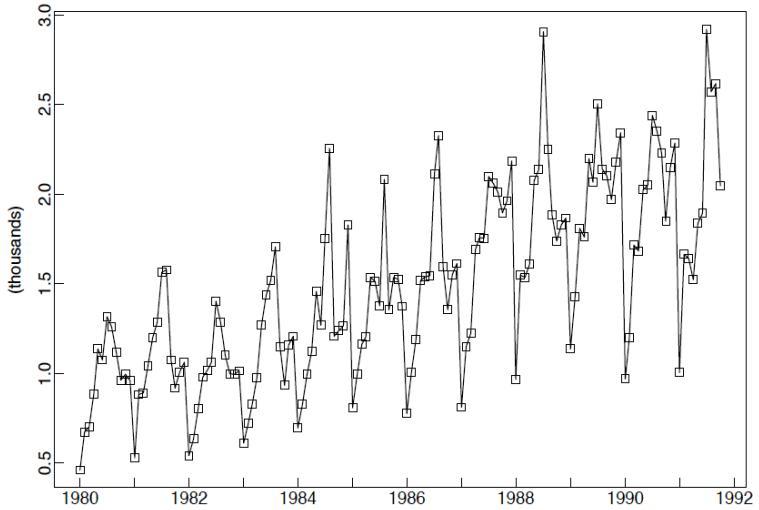
A **time series** is a set of observations x_t , each one being recorded at a specific time t .

A **discrete-time time series** is one in which the set T_0 of times at which observations are made is a discrete set, e.g. when observations are made at fixed time intervals.

Continuous-time time series are obtained when observations are recorded continuously over some time interval, e.g., when $T_0 = [0,1]$.

Introduction to time series analysis

The Australian redwine sales, Jan.‘80–Oct.‘91 (Brockwell & Davis)
In this case the set T_0 consists of the 142 times {(Jan. 1980), (Feb. 1980), ...,(Oct. 1991)}.



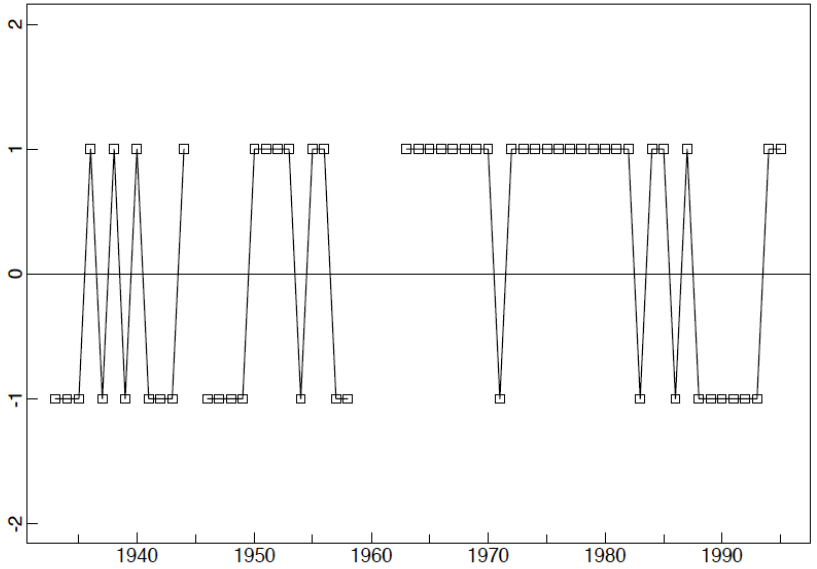
Given a set of n observations made at uniformly spaced time intervals, it is often convenient to rescale the time axis in such a way that T_0 becomes the set of integers {1,2,...,n}.
In the present example this amounts to measuring time in months with (Jan.1980) as month1. Then T_0 is the set {1,2,...,142}.

It appears from the graph that the sales have an upward **trend** and a **seasonal pattern** with a peak in July and a trough in January.

Introduction to time series analysis

Results of the all-star baseball games,1933–1995, by plotting, where

$$x_t = \begin{cases} 1 & \text{if the National League won in year } t, \\ -1 & \text{if the American League won in year } t. \end{cases}$$

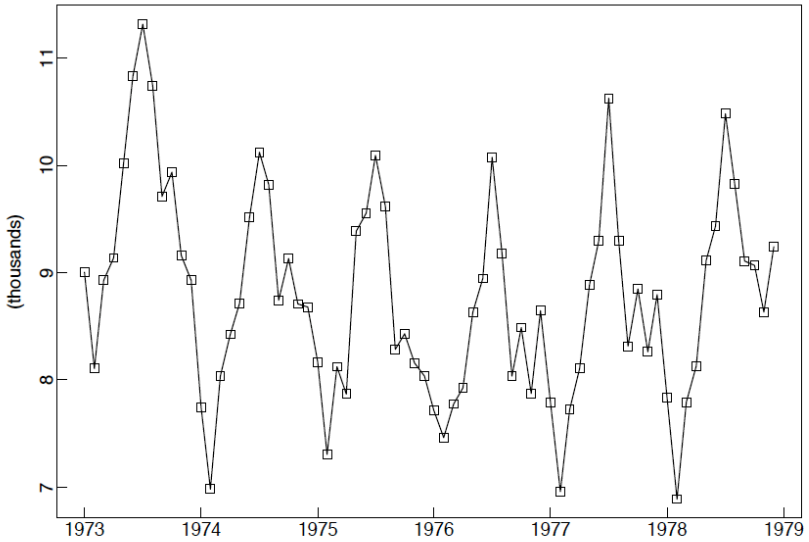


This is a series with only two possible values, ± 1.

It also has some missing values, since no game was played in1945, and two games were scheduled for each of the years 1959–1962.

Introduction to time series analysis

The monthly accidental deaths data,1973–1978, in USA.



Like the red wine sales, the monthly accidental death figures show a strong **seasonal pattern**, with the maximum for each year occurring in July and the minimum for each year occurring in February.

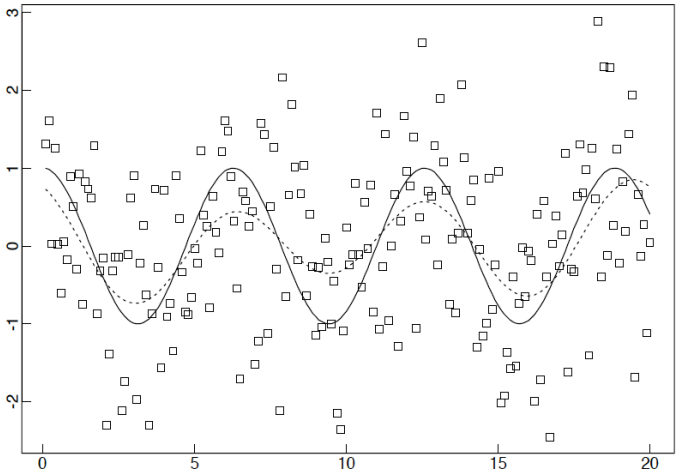
The presence of a **trend** in this figure is much less apparent than in the wine sales. We shall later consider the problem of representing the data as the sum of a trend, a seasonal component, and a residual term.

Introduction to time series analysis

The series $\{X\}$ of

$$X_t = \cos\left(\frac{t}{10}\right) + N_t, \quad t = 1, 2, \ldots, 200,$$

where $\{N_t\}$ is a sequence of independent normal random variables, with mean 0 and variance 0.25. Such a series is often referred to as **signal plus noise**, the signal being the smooth function, S_t = cos(t / 10) in this case.



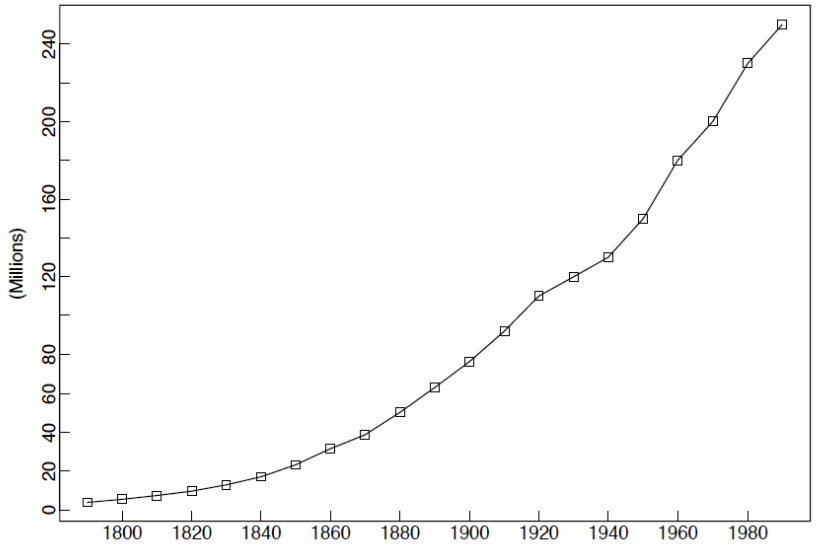
Given only the data X_t, how can we determine the unknown signal component? There are many approaches to this general problem under varying assumptions about the signal and the noise. One simple approach is to smooth the data by expressing X_t as a sum of sine waves of various frequencies and eliminating the high-frequency components.

If we do this to the values of $\{X_t\}$ shown in the figure and retain only the lowest 3.5% of the frequency components, we obtain the estimate of the signal also shown in the figure.

The waveform of the signal is quite close to that of the true signal in this case, although its amplitude is somewhat smaller.

Introduction to time series analysis

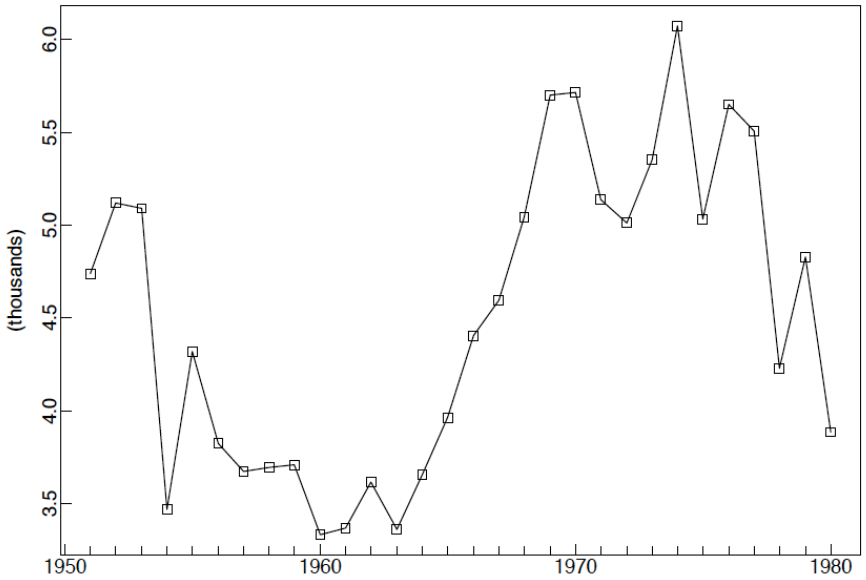
Population of the USA at ten-year intervals,1790–1990.



The graph suggests the possibility of fitting a quadratic or exponential trend to the data.

Introduction to time series analysis

Strikes in the USA.,1951–1980.



The number of strikes appears to fluctuate erratically about a slowly changing level.

Objectives of time series analysis

An important part of the analysis of a time series is the selection of a suitable probability model (or class of models) for the data.

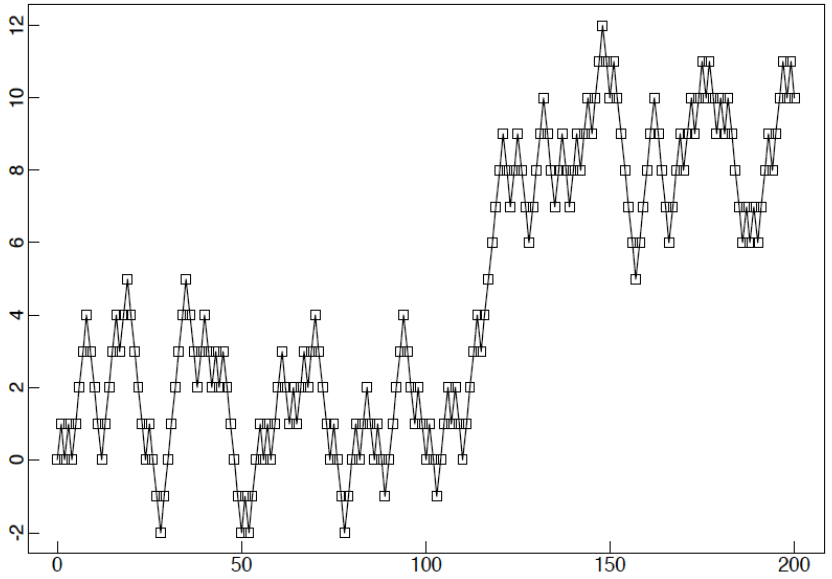
To allow for the possibly unpredictable nature of future observations it is natural to suppose that each observation x_t is a realized value of a certain random variable X_t .

Definition: A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

Remark. We shall frequently use the term time series to mean both the data and the process of which it is a realization.

Objectives of time series analysis

The figure shows one of many possible realizations of $\{S_t\ , t = 1, \ldots, 200\}$, where $\{S_t\}$ is a sequence of random variables.



In most practical problems involving time series we see only one realization.

For example, there is only one available realization of Fort Collins's annual rainfall for the years 1900–1996, but we imagine it to be one of the many sequences that might have occurred.

Objectives of time series analysis

A **complete probabilistic time series model** for the sequence of random variables $\{X_1, X_2, \ldots\}$ would specify all of the joint distributions of the random vectors $(X_1, \ldots, X_n)'$, n = 1, 2, . . ., or equivalently all of the probabilities

$$P[X_1 \leq x_1, \ldots, X_n \leq x_n], \quad -\infty < x_1, \ldots, x_n < \infty, \quad n = 1, 2, \ldots.$$

Such a specification is rarely used in time series analysis, since in general it will contain far too many parameters to be estimated from the available data.

Instead we specify only the **first- and second-order moments** of the joint distributions, i.e., the expected values EX_t and the expected products $E(X_{t+h}X_t)$, t = 1, 2, . . ., h = 0, 1, 2, . . ., focusing on properties of the sequence $\{X_t\}$ that depend only on these.

Objectives of time series analysis

Such properties of $\{X_t\}$ are referred to as second-order properties.

In the particular case where all the joint distributions are multivariate normal, the second-order properties of $\{X_t\}$ completely determine the joint distributions and hence give a complete probabilistic characterization of the sequence.

In general we shall loose a certain amount of information by looking at time series "through second-order spectacles".

As we shall see later, the theory of minimum mean squared error linear prediction depends only on the second-order properties, thus providing further justification for the use of the second-order characterization of time series models.

For V3

For next week:

Read paper „Effects of insufficient sleep …“

Only Intro + Methods section

In V3: results and discussion