RESEARCH REPORTS
Biological

Y. Gafni[1], A.A. Ptitsyn[2,3], Y. Zilberman[1], G. Pelled[1], J.M. Gimble[3], and D. Gazit[1,4*]

[1]Hebrew University– Hadassah Medical Campus, PO BOX 12272, Ein Kerem, Jerusalem 91120, Israel; [2]Colorado State University, Ft. Collins, CO, USA; [3]Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, LA, USA; and [4]Stem Cell Therapeutics Research Lab—Department of Surgery, Cedars Sinai Medical Center, Los Angeles, CA, USA; *corresponding author, dgaz@cc.huji.ac.il

J Dent Res 88(1):45-50, January, 2009

Circadian Rhythm of Osteocalcin in the Maxillomandibular Complex

APPENDIX

METHODS

Animal Studies

Data Processing and Time Series Analysis

Intact mice of both genders, transgenic for the human osteocalcin promoter/luciferase reporter gene, were used at ages 1, 3, and 5 mos and 1.5 yrs (N = 5 for each gender and time-point). Mice were trained for a 12-hour light/dark cycle for a minimum of 2 wks prior to each study. Each animal was sequentially analyzed at times 0, 6, 12, 18, and 24 hrs of the light/dark cycle throughout a single 24-hour period. Data were analyzed statistically as detailed below.

Quantitative analysis of luciferase expression was performed with the MetaImaging series 4.6 software (Molecular Devices, Downingtown, PA, USA), with a constant measurement field for all time-points. Results are presented in integrated luciferase units.

Seven skeletal sites were analyzed: calvaria, tail, maxillomandibular complex, carpals, and tarsals.

In vivo Bioluminescence Imaging

In all experiments, the animals were anesthetized with a mixture of ketamine and xylazine, injected intraperitoneally at 1 μL/g body weight. Ten min before the light emissions were monitored, each animal was given an intraperitoneal injection of beetle luciferin (Promega Corp., Madison, WI, USA) in phosphate-buffered saline (PBS) at 126 mg/kg body weight. The mice were placed in a light-tight chamber (a dark box), and a gray-scale image of the animal was first recorded in dimmed light. Photon emission was then integrated over a period of 2 min and recorded as pseudocolor images (Contag *et al.*, 1997; Honigman *et al.*, 2001; Iris *et al.*, 2003). Based on our published data, an internal control (at the tail level) was needed for quantification of the results, because our hOC-Luc mice differed in their luciferase expression levels (Iris *et al.*, 2003).

The imaging unit consisted of an intensified Peltier cooled charge-coupled device (CCCD), model LN/CCD-1300EB, which was equipped with an ST-133 controller and a 50-mm Nikon lens (Roper Scientific, Princeton Instrument, Trenton, NJ, USA). In this system, a pseudocolor image represents light intensity (with blue representing least intense and red most intense). In all cases, the integrated light is the result of a two-minute exposure and acquisition. The exposure conditions (including time, f/stop, position of the stage, binding ratio, and time after injection with luciferin) were maintained at identical levels so that all measurements would be comparable.

Computational Methods

Phase Assignment

Consider a series of microarray expression values for gene *x,* with *N* samples of the form $Y = \{x_0, x_1, x_2, \ldots x_{N-1}\}$. We assigned phase to each expression time series by computing cross-correlation

$$R(f) = \frac{\sum_0^{N-1} (x_i - \bar{x})(y_f - \bar{y})}{\sum_0^{N-1} (x_i - \bar{x})(y_i - \bar{y})}$$

where x is a gene expression time series of N points and y is an artificially generated profile of ideal cosine function

$$y_i = \cos\left(\frac{2\pi}{p} * i\right)$$

where p is the number of time-points in a complete circadian cycle, *i.e.*, 4 in this dataset. To account for all phases, the artificial cosine curve profile has been regenerated with a phase shift by 1 time-point. The highest correlation among all possible phase shifts was assigned as the most probable phase. The significance of periodicity was not assessed before the phase assignment. Consequently, we assigned the most probable potential phase to all profiles, both truly periodic and chaotic.

Spectral Analysis

The time series $Y = \{x_0, x_1, x_2, \ldots x_{N-1}\}$ can be converted from time-domain, where each variable represents a measurement in time to a frequency domain by the Discrete Fourier Transform (DFT) algorithm. Frequency domain representation of the series of experiments is also known as a periodogram, which can be denoted by $I(\omega)$:

$$I(\omega) = \frac{1}{N}\left|\sum_{t=0}^{N-1} x_t e^{(-i\omega t)}\right|^2, \omega \in [0, \pi].$$

If a time series has a significant sinusoidal component with frequency $\omega \in [0, \pi]$, then the periodogram exhibits a peak at that frequency with a high probability. Conversely, if the time series is a purely random process (a.k.a. "white noise"), then the plot of the periodogram against the Fourier frequencies approaches a straight line (Priestley, 1981).

Fisher's *g* Test

Significance of the observed periodicity can be estimated by Fisher *g* statistics, as recently recommended (Wichert *et al.,* 2004). Fisher derived an exact test of the maximum periodogram coordinate by introducing the *g* statistic

$$g = \frac{\max_k I(\omega_k)}{\sum_{k=1}^{N/2} I(\omega_k)},$$

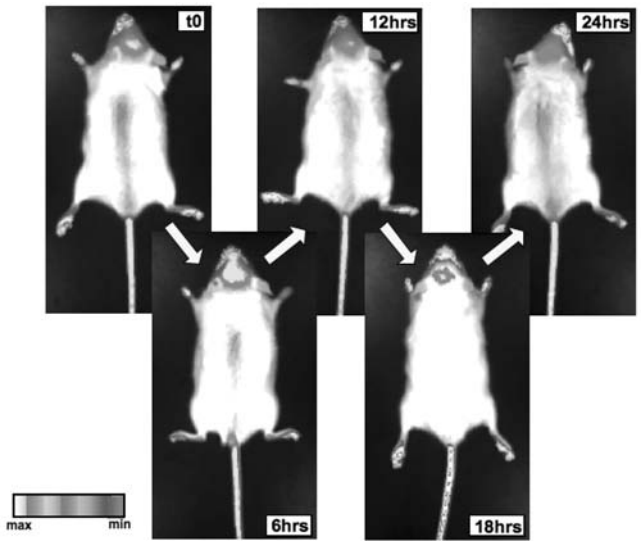
where $I(\omega_k)$ is a *k*-th peak of the periodogram. Large values of *g* indicate a non-random periodicity. To calculate the *p*-value of the test under the null hypothesis, we used the exact distribution of *g* given by

$$P(g > x) = \sum_{p=1}^{1/x}\left[(-1)^p \frac{n!}{p!(n-p)!}(1-px)^{n-1}\right],$$

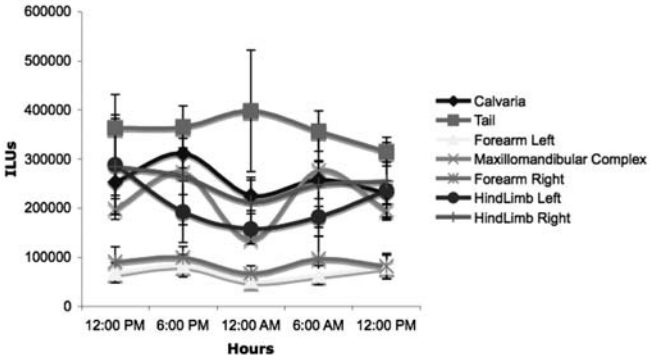
where $n = [N/2]$ and p is the largest integer less than $1/x$. This algorithm closely follows the guidelines recommended for analysis of periodicities in time-series microarray data (Wichert *et al.*), with the exception that we applied our own C++ code instead of R scripts.

Permutation Test

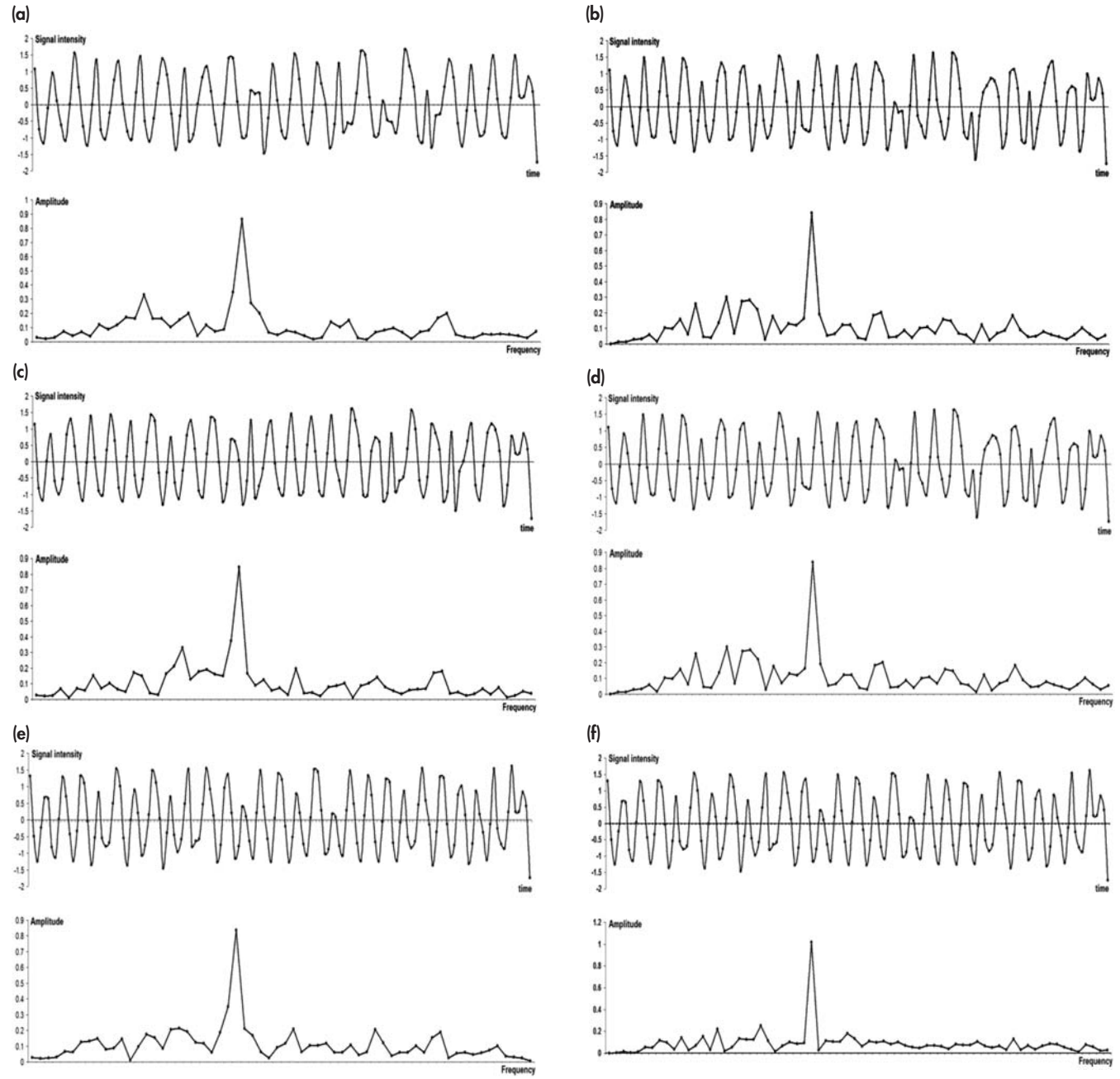
The alternative test for significance of a particular (in our case circadian) periodicity among large numbers of gene expression profiles is based on the random permutation procedure. Consider a time series $Y = \{x_0, x_1, x_2, \ldots x_{N-1}\}$, in which technical variation approaches or even exceeds the amplitude of periodic expression. In a very short time series, there is a significant probability of observing a periodicity due to stochastic reasons. However, the periodic change of the base expression level can still be identified in spite of the high noise level. Let *YR* be a random permutation of the time series *Y* and its corresponding periodogram *IR(w)*. If the periodogram *IY(w)* contains a significant peak corresponding to a particular frequency, this peak results from a particular order of observation is in the *Y*. A random permutation would preserve the same noise level, but not the periodicity. After DFT, a periodogram *IR(w)* represents only the peaks occurring by chance. To avoid random re-institution of periodicity of length *T* (in this case circadian), we generate *YR* by multiple shuffling of randomly selected time-points $x_n \alpha x_m$, where $T \nmid |n - m|$, *i.e.,* each shuffle swaps time-points from different



Appendix Figure 1. *In vivo* whole-body bioluminescence imaging of transgenic mouse at 5 time-points within a 24-hour cycle of a representative animal. The mouse displayed is a 5-month-old female. Intact mice of both genders for the human osteocalcin promoter/luciferase reporter gene were used at ages 1, 3, and 5 mos and 1.5 yrs. N = 5 for each gender and time-point. Each animal was sequentially analyzed at t = 0, 6, 12, 18, and 24 hrs of the light/dark cycle throughout one 24-hour period for each animal.



Appendix Figure 2. Quantified *in vivo* bioluminescence imaging of a transgenic mouse at 5 time-points within a 24-hour cycle. Chart shows a representative group of 5-month-old females. Results compare the osteocalcin circadian patterns of expression for individual data sites with the maxillomandibular complex. Results show no statistical significance within each individual skeletal site, except for the maxillomandibular complex. Calvaria, left carpals, right carpals, left tarsals, right tarsals, and tail.



Appendix Figure 3. Concatemer of data for individual data sites. Circadian oscillation of osteocalcin promoter/luciferase reporter in maxillo-mandibular complex of transgenic mice *in vivo*. Anesthetized mice (n = 23-28) were examined in a bioimaging device for light emission at six-hour intervals over a 24-hour period immediately following injection with luciferin. Data from the individual animals have been concatenated in order of emission amplitude and frequency profile (upper panel). Periodogram based on a common scale of magnitude (lower panel). (**a**) Calvaria, (**b**) left carpals, (**c**) right carpals, (**d**) left tarsals, (**e**) right tarsals, and (**f**) tail.

phases. Comparing permutations with deliberately wiped out periodicity with the original time series, we can estimate whether a particular order of observations (*i.e.*, time series) is important. For each gene expression profile, we generate two series of *min(n!,1000)* random permutations. Each permutated series *YR* is transformed to the frequency domain, and a single peak of the periodogram *IR(w)* is stored. The *p*-value for the null hypothesis of the random nature of a particular peak of a periodogram can be estimated by comparing the stored *IR(w)* values with the observed *I(w)*:

$$p = \frac{K_{IR(\omega) \geq IY(\omega)}}{\min(n!, 1000)} \quad .$$

Here, *K* is the number of permutated series *YR* for which the circadian peak of a periodogram is higher or equal to that of the

original time series *Y*. A high *p*-value exceeding the threshold—for example, 0.05—means that at least 5 out of 100 random permutations of the time series produce a periodogram with the same or higher peak, corresponding to a given periodicity. Low *p*-values indicate a significant difference between periodograms *IR(w),* preserving circadian periodicity and purely random periodograms with the same level of technical variation.

Autocorrelation

For a given a discrete time series $Y = \{x_0, x_1, x_2, \ldots x_{N-1}\}$, the autocorrelation is simply the correlation of the expression profile against itself with a frame shift of k datapoints (where, $0 \le k \le N-1$, often referred to as the lag).

For the time shift f, defined as $f = i + k$ if $i + k < N$ and $f = i + k - N$ otherwise.

$$R(f) = \frac{\sum_0^{N-1} (x_i - \bar{x})(x_f - \bar{x})}{\sum_0^{N-1} (x_i - \bar{x})^2}.$$

For each time series, we calculate the maximum positive *R(f)* among all possible phase shifts *f* and use tabulated 0.05 significance cut-off values for correlation coefficient. Time series that shows significant autocorrelation *R(f)* with the lag *f* corresponding to one day (4 datapoints) are considered to be circadially expressed.

Data Analysis Pipeline

The data for computational analysis were prepared in the form of a text (tab-delimited) file containing time series in each row. Samples from different bones were analyzed separately. All absolute values x of signal intensity in a time series Y are converted to z-score:

$$z_i = \frac{\bar{x} - x_i}{\sigma_Y},$$

where σ_Y is a standard deviation and $\bar{x}$ is a mean or *mesor* value of time series *Y*. The first step of analysis assigns the most likely phase to each of the individual timelines. Profiles are then separated into groups with the same assigned phase, and each phase group is analyzed separately. Within the same-phase group, expression profiles have been smoothed by the Savitsky-Golay polynomial algorithm (Savitzky & Golay, 1964). Periodicity of expression profiles within same-phase groups is tested by autocorrelation, Fisher's *g* test (Fisher, 1929), Pt-test (Ptitsyn *et al.,* 2006), and the Kolmogorov-Smirnov fit of permutated periodogram (Ptitsyn *et al.,* 2006).

Challenges of Data Analysis and Interpretation (Discussion)

Like other studies of circadian gene expression, the major challenge for analysis of periodicity is associated with an extremely low sampling rate. While, in microarray studies, increasing the number of time-points at which gene expression is measured is prohibitively expensive, the invasive procedure collecting the data from animals in this study makes shorter spacing between time-points impossible. As in microarray data, at each time-point the level of gene expression is estimated with a large degree of stochastic variation, which creates a high level of noise interfering with the rhythmic signal. As in microarray experiments, for the pattern analysis, we consider these biological replicates as a replication of a complete period rather than single points. Since our goal is to detect a pattern rather than to improve the precision of gene expression estimation at each time-point, it is correct to concatenate the replicated profiles, producing a single extended expression profile. Most of the microarray studies of circadian gene expression are limited by the length of the timeline, covering not more than 2 complete daily periods. In this study, each group of profiles contains 5 or 6 biological replicates, which improves our chances of identifying the periodicity compared with the microarray data. In contrast, a lower sampling rate created an additional challenge. While the general pattern remains similar, repeating every 24 hrs, stochastic variation of expression levels may lead to variations in the time of peak expression. If the acrophase (time of the peak value) is different in one of 5 or 6 periods of oscillation, the entire series will fail the statistical test for periodicity. With only 4 points *per* period, even a small random deviation of measured expression level may lead to incorrect estimation of acrophase and discrepancy between visually obvious repeating patterns and statistics. To overcome this contradiction, we decided to perform phase classification of expression profiles before testing these profiles for periodicity. This strategy does not interfere with the statistical test for periodicity. The potentially most probable phase can be assigned to a chaotic non-periodic expression profile. However, after we grouped expression profiles together by phase, we found only a small fraction of expression profiles showing chaotic variation, with no detectable circadian periodicity.

APPENDIX REFERENCES

Contag CH, Spilman SD, Contag PR, Oshiro M, Eames B, Dennery P, *et al.* (1997). Visualizing gene expression in living mammals using a bioluminescent reporter. *Photochem Photobiol* 66:523-531.

Fisher R (1929). Tests of significance in harmonic analysis. *Proc R Soc B* 125:54-59.

Honigman A, Zeira E, Ohana P, Abramovitz R, Tavor E, Bar I, *et al.* (2001). Imaging transgene expression in live animals. *Mol Ther* 4:239-249.

Iris B, Zilberman Y, Zeira E, Galun E, Honigman A, Turgeman G, *et al.* (2003). Molecular imaging of the skeleton: quantitative real-time bioluminescence monitoring gene expression in bone repair and development. *J Bone Miner Res* 18:570-578.

Priestley MB (1981). Spectral analysis and time series. London: Academic Press, pp. 816-890.

Ptitsyn AA, Zvonic S, Gimble JM (2006). Permutation test for periodicity in short time series data. *BMC Bioinformatics* 7(Suppl2):S10.

Savitzky A, Golay MJE (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627-1639.

Wichert S, Fokianos K, Strimmer K (2004). Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* 20:5-20.