Referee: 1

This manuscript develops ZeitZeiger, a method and software for predicting the time-of-day from high-throughput biological measurements, such as gene expression measurements. The method first learns a sparse representation of the variation associated with the periodic variable in the training set (using Sparse Principal Component Analysis) and then uses maximum likelihood estimation to predict the time for a test observation. ZeitZeiger is trained using a comprehensive dataset of circadian measurements taken across multiple organs. Using 13 core clock genes, ZeitZeiger yields a multi-organ circadian time predictor and compares favorably in terms of both speed and accuracy with one alternative program.

Overall the topic of the article is interesting and reasonably appropriate for NAR, the methods are sound, and the resulting program could be useful for circadian analyses. The literature references are appropriate and sufficient. Thus in my opinion the basic results are publishable. However I found the manuscript *very poorly written* and it should be significantly revised before it can be considered for publication.

Thanks for your feedback. We have extensively revised the manuscript and detail the specific changes below. Because we have made so many changes to the text (but few to the content), only major changes are highlighted in yellow.

1) First the manuscript is too long and should be shortened to about half of its current length. The manuscript is very verbose and repetitive (some examples are given below).

We have trimmed the manuscript by about 20%; not quite half, but considerable nonetheless!

2) A good example is the Introduction which could be reduced to a single short paragraph essentially saying that the goal of the manuscript is to predict circadian time from microarray measurements. The current introduction is too long and too elementary for a scientific article in NAR. It is more geared towards a review article for non-experts.

We have condensed the Introduction and made it less elementary. We emphasize, however, that the goal of the manuscript is to present a general method for regularized supervised learning on oscillatory data. Predicting circadian time from microarray measurements is simply the example we use to demonstrate the method.

3) There is a lot of verbiage throughout the manuscript trying to hint that this is the first time that supervised machine learning methods are applied to periodic signals. This is far from the truth as there is a substantial machine learning literature (both supervised and unsupervised) on periodic signals--all such claims should be removed as they simply reduce the credibility of the paper.

We have removed all such implications.

4) Similarly the claim that this is " the first multi-organ predictor of circadian time" seems unnecessary. The paper can stand on its own technical contributions and does not need to make

such a weak claim. (The claim is technically correct but only because of the qualifier "multi-organ". There is already at least one predictor of circadian time.)

We have removed all language related to "the first" multi-organ predictor.

5) The following sentences "the output variable in an oscillatory system is periodic, i.e., it is a closed loop" or "a periodic variable is a closed loop" are completely devoid of any scientific value. What does "closed loop" mean and what does it bring scientifically that is not already contained in the term "periodic"? To make matters worse, these sentences are repeated more or less verbatim multiple times (I have marked 7 occurrences, but there may be more).

We have removed all language related to "closed loop".

6) There are many occurrences of sentences that are almost completely disconnected from the previous text. This should be avoided. One example, out of many, is the last sentence of the paragraph starting with "The state of the circadian oscillator at any particular time....".

We have removed the offending sentence and revised or removed many others.

7) Several technical details are unclear. For instance, is the predictor animal specific? How does it deal with the fact that some animals (e.g. mice) are nocturnal?

We have added text making it explicit that the multi-organ predictor is specific to mice. We have only tested the predictor on data from mice, and make no claims about whether it would work in other organisms. Cross-species comparison of circadian gene expression is an interesting topic for future investigation, but we feel it is beyond the scope of the current manuscript.

8) Training data for ZeitZeiger should be a matrix X ..... What are the values of n and p? What are the "features"? This is not clearly explained. Do you need fine-grained time resolution in the training set?

We have clarified these points.

ZeitZeiger is a general method for supervised learning, like the lasso or random forests, so it doesn't require specific values of n and p. By "features", we mean the variables that were measured for each observation, which could be genes, metabolites, or anything else that could be used to make a prediction. As far as the values of n and p for the multi-organ predictor, the training data consisted of 21115 genes (p) measured in 288 samples (n).

The need for high time resolution will likely vary from one system to another (and even within the same system, from one feature to another). That said, we have now performed a comprehensive analysis of the effect of time resolution and number of samples in the training set (now Supplementary Figure S5). With the caveat that our results are from only one dataset (GSE54650) measured on one biological system, it appears that ZeitZeiger does not need fine-grained time resolution. In fact, ZeitZeiger achieved a median absolute error (on test samples) of

about 1 h, based on a training set of 12 samples from 3 time-points. On the same training and test samples, the molecular-timetable method showed a median absolute error of about 4 h.

9) It is not clear what must be fed to ZeitZeiger in prediction mode? Is it the level of expression of the core genes measure at one time point? Is the information contained in the other genes discarded? Is there a different predictor for each organ? Or is the organ part of the input to the predictor?

We have clarified these points in the text.

In prediction mode, ZeitZeiger uses the measurements of the features that contribute to the SPCs. So yes, for the multi-organ predictor, ZeitZeiger uses the expression of the selected genes in one sample (which is from one time-point). All other genes are ignored.

Regardless of organ, the only input to the predictor is the expression of the selected genes. ZeitZeiger was given no information about which samples came from which organ, other than to make the folds for cross-validation.

10) What about missing time points and missing measurements in the trianing or test data? How are these handled?

We have clarified this in the Methods. Training data can have missing measurements; observations with missing measurements for a particular feature are ignored during that feature's spline fitting (but are used for the other features' spline fits). Test data cannot have missing measurements for the features that are used in the predictor (but these will be typically be only a small subset of all the features). There are no restrictions on the number or spacing of time points, so the time points do not have to be evenly spaced and there could be a different number of replicates for different times.

11) "Because time is periodic here, ZeitZeiger can compute the likelihood for all possible times, so there is no worry of getting stuck in a local minimum". By itself this sentence is cryptic. I assume what the authors are doing is coarsely discretizing time (e.g. 12 or 24 points, and computing the likelihood at these few points?)

We apologize for the ambiguity. We have now clarified in the Methods how ZeitZeiger finds the time of maximum likelihood. In short, ZeitZeiger does not simply calculate the likelihood at a set number of points, but instead uses bound-constrained optimization to find the time at which the likelihood is maximal.

12) ZeitZeiger is described in the Methods, and then once again at the beginning of the Results!!!!

We have condensed the description of ZeitZeiger in the Results section to just a couple sentences, for those readers who do not wish to read the full technical description in the Methods.

Referee: 2

Summary:

This paper presents a new approach to identifying genes that show an oscillatory pattern. The approach consists of fitting a periodic cubic spline to each gene, then standardizes the model fits for each gene and calculates sparse principal components of the resulting model fits, then each SPC is assumed to be a smooth function of time with normally distributed errors, and a maximum likelihood model to estimate the time for any test observation after projecting it onto the sparse PCs. They apply this approach to a range of data sets and benchmark the approach against the molecular-timetable work previously described by Ueda et al. Overall the paper is nicely written, the code appears to be thoroughly documented, and the approach is interesting. I have honestly only have two minor comments, but otherwise enjoyed the paper and thought it was really well done.

Thank you very much.

Minor Comments:

(1) The code is nicely reproducible, but the impact/use of the method would be dramatically increased if the authors made the effort to create a complete R package for their analysis and submit it to CRAN/Bioconductor.

You're right. We have now created an R package for ZeitZeiger, available on GitHub (https://github.com/jakejh/zeitzeiger). We intend to submit it to Bioconductor once we further refine the documentation. While the R package will enable new analyses, we will still publish the complete data and results from this manuscript on Zenodo.

(2) It seems like there are a few other methods that could be used to benchmark against, including the oscope approach that was very recently published http://www.nature.com/nmeth/journal/v12/n10/full/nmeth.3549.html, which has a corresponding Bioconductor package: https://bioconductor.org/packages/release/bioc/html/Oscope.html so the comparison should be straightforward. The Oscope method may work well since it is also consistent with the observation on page 10 that the circadian clock may be approximated by a two dimensional oscillator.

Thanks for the suggestion. We have tried analyzing our data with Oscope, and we do cite it in the manuscript. However, we don't believe that Oscope can be directly compared to ZeitZeiger, because Oscope is unsupervised and currently has no way to use the sample labels (e.g., the value of circadian time). Thus, although Oscope can learn an unsupervised ordering of the samples, it is not clear how to combine that learned order with the true labels to make a prediction. The same is true of another recently published method called Cycler (http://www.nature.com/nmeth/journal/v12/n10/full/nmeth.3545.html). Another set of methods called Cyclone (https://github.com/PMBio/cyclone) was just published for predicting cell-cycle stage from single-cell RNA-seq data, but these methods predict a discrete class, not a continuous value, so they are not comparable to ZeitZeiger either.

Referee: 2

Comments for the Author
The authors have addressed my comments.

Referee: 1

Comments for the Author
The authors have addressed many of the concerns of the reviewers and incorporated many of their suggestions. As a result, the manuscript has been considerably improved.

My only residual concern is with the length of the manuscript. It has been tightened and reads much better but I think it could be shortened even further.
For example, the first paragraph of the Results section summarizing the methods seems superfluous. The authors state
"We have condensed the description of ZeitZeiger in the Results section to just a couple sentences, for those readers who do not wish to read the full technical description in the Methods. "
However the paragraph by itself seems fairly cryptic and it will not really help readers who do not wish to read the Methods. These readers will have already acquired a general sense for the methods used in the manuscript from the Abstract and the Introduction. Thus I would delete it.
Likewise, I would further shorten the rest of the manuscript by removing any repetitions or statements that are not essential. Some material can always be added to the Supp. Mat.

Thanks again for your feedback. It's great to hear that you think that manuscript reads much better.

As you suggested, we have removed the first paragraph in the Results section. We have also scoured the manuscript for other instances of superfluity, shortening the manuscript by an additional 5% since our previous submission. As we have already moved much material to the supplement, we believe we are now in the area of diminishing returns.