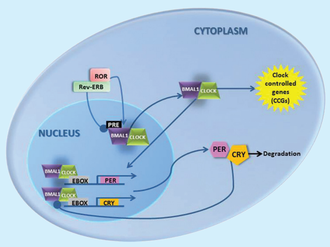
NITECAP : A novel method and interface for the identification of circadian behavior in highly parallel time-course data

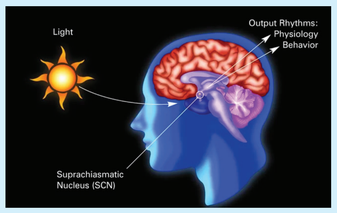
Thomas G. Brooks*, Cris W. Lawrence, Nicholas F. Lahens, Soumyashant Nayak, Dimitra Sarantopoulou, Garret A. FitzGerald, Gregory R. Grant
The University of Pennsylvania, Institute for Translational Medicine and Therapeutics, Dept. of Genetics, Dept. of Pharmacology

Contact: thobr@sas.upenn.edu

The Molecular Clock

- Every cell has a molecular clock.
  - A transcriptional/translational feedback loop that takes roughly 24 hours to complete.
  - Key Genes in mammals: BMAL1, CLOCK, PER1/2, CRY1/2
  - You can break the clock by knocking out BMAL1
- Within each tissue, the cellular clocks are coordinated
- There is a central clock located in the SCN brain region.
- Clocks cycle indefinitely even in total darkness.
- Clocks can be entrained by environmental factors such as light, feeding, temperature, etc.





Importance

- Relevant to disease, pharmacology, personalized medicine.
  - E.g. some statins are only effective if taken in the evening.
- And relevant to daily life by way of jet-lag, shift work, etc.
  - Extended exposure is associated with adverse health.

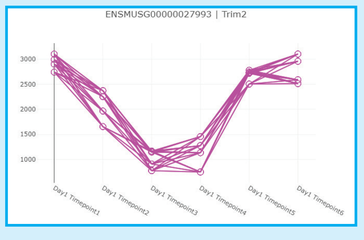
- Circadian biology is gaining wide appreciation.
  - Yet there is a lack of tools to perform circadian analysis, especially in "big data".

Terminology

- ***Feature*** : any biological entity that can be measured. E.g. RNA, protein, metabolite, etc...
- ***Period*** : a span of time under which features behave cyclically.

The Data

- Time course profiles of features.
  - Potentially thousands, or tens of thousands, of features in parallel.
- Measurements taken at regular intervals across the period.



Profile of one Feature

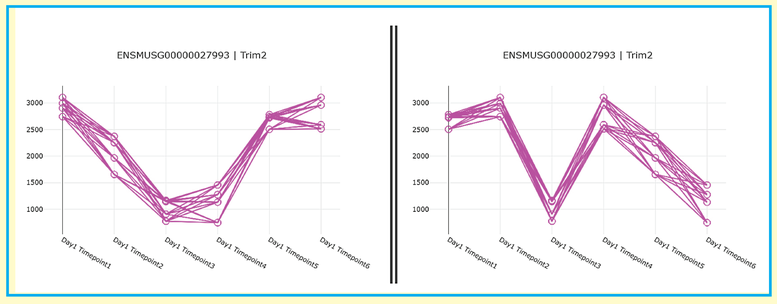
The Problem

- Given profiles for a large number of features, find those that present *significant* circadian behavior.
- There are several algorithms available for this problem.
  - We have developed a permutation based method called NITECAP (described later).
  - **N**on-parametric **I**dentification **T**ool **E**nabling **C**ircadian **A**nalysis in **P**arallel

- **Difficulty for all methods**: find an appropriate significance cutoff to determine a significant set of features with which to proceed.
- There is often no clear division between what is and what is not circadian.
  - Therefore a choice of cutoff is inevitably subjective
- Visual inspection facilitates this judgment call.
- *However, up to now there has been no easy way of quickly scanning the profiles.*
  - Nor of comparing profiles of the same feature in two or more data sets.

NITECAP Algorithm

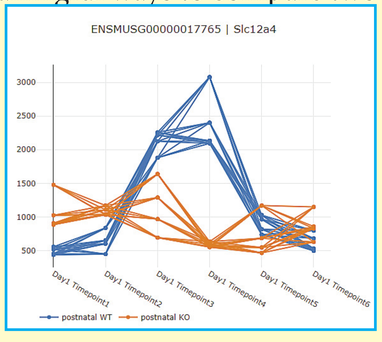
- There are multiple possible definitions of which profiles should be called "circadian"
- NITECAP exploits the ordering of time points with a permutation approach.
  - **The concept:** Consecutive time points should be more similar than two randomly chosen time points in circadian features.
  - **The statistic:** Take the difference between data points *in adjacent time points*
    - Then sum these differences over all such pairs of points.
  - **The test:** If the profile is circadian then randomly permuting the time points should tend to increase the statistic.



Regular Profile Permuted Profile

Two Sample Comparisons

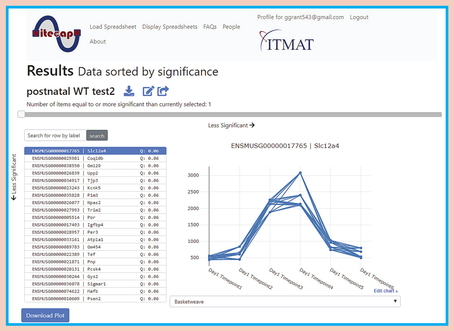
- There are potentially many meaningful ways to compare two profiles.
- Two methods are currently implemented:
  1. Dampening analysis (with a permutation test for significance)
  2. Two-way ANOVA



https://nitecap.org

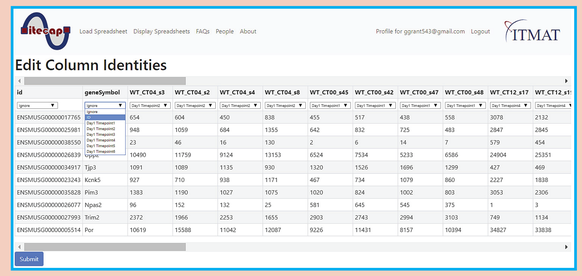
Interface

1. ***Allows users to quickly scan through profiles to determine an appropriate significance threshold.***



Use the horizontal slider bar to quickly flash through the profiles, to make sure the chosen significance threshold agrees with intuition.

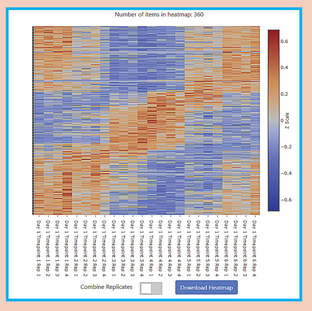
2. Upload and configure spreadsheets.



Configuration page, here you identify which columns correspond to which time points, and which columns give feature IDs.

3. Sort profiles by various statistics
   - Nitecap, JTK, ANOVA, more to come...

4. Draw heatmaps and PCA plots.
   - Once a significance threshold is chosen, a heatmap is drawn for the significant features, sorted by phase.



5. Create user accounts and save your data.
6. Share views of your data by emailing a link.
7. Compare two data sets for differential effects.