Exploring the Viability of a PLSR-based Machine Learning Method for Predicting Circadian Phase in Cancer Patients

Elizabeth Epstein, Karen Reyes, Evan Fagan, Fritz Huggins
Mentor: Jonathan Trinidad
Advisor: Casey Diekman
Department of Mathematical Sciences
New Jersey Institute of Technology, Newark NJ 07102

The circadian clock is an internal biological timekeeping system that regulates daily patterns in physiology and behavior, such as an individual's sleep-wake cycle. The molecular mechanism underlying circadian rhythm generation is an intracellular clock composed of transcriptional and translational feedback loops, which leads to 24-hour rhythms in gene expression. The tendency for gene expression to peak at specific times of the day has implications in cancer therapy, as the effect of such treatment is tied to the manner in which genes are expressed at the time of administration. However, because there is significant variation in the phase of circadian rhythms across individuals, the practical application of circadian-based chronotherapy requires an efficient and cost-effective means through which a person's unique circadian phase can be measured. The currently accepted system used to assess an individual's circadian rhythm is a tedious and expensive process that necessitates gathering several blood samples from an individual over a relatively short period of time. While effective, this system is not practical in a clinical setting.

This project aims to explore a recently developed method for assessing an individual's circadian phase more efficiently. While a number of methods of varying levels of accuracy have been developed, most require a minimum of two samples to be taken from an individual. The Partial Least Squares Regression (PLSR) method developed by Laing et al. (2017) stands out in the fact that it requires only a single blood draw. The PLSR method uses a machine learning model to extrapolate circadian time based on an assay of about 26,000 genes. The model in its current form has only been validated for use in blood samples from healthy individuals and has never been tested on cancer patients.

We are gaining familiarity with Laing et al.'s method by first applying it to a new dataset from adipose tissue that was not part of the training set for the machine learning algorithm. Next, we will apply the method to publicly available gene expression datasets from tumor samples to assess how well the algorithm can predict circadian phase in cancerous tissues.

Laing, Emma E, et al. "Blood Transcriptome Based Biomarkers for Human Circadian Phase." *ELife*, vol. 6, 2017, doi:10.7554/elife.20214.