Title: Detecting Rhythmic Components transcript and the effect of 2,3,7,8-Tetrachlorodibenzo-ρ-dioxin in hepatic metabolic activity in mice

Marri Kwabena Daniel

Project advisor: Sudin Bhattacharya

Abstract

Aryl hydrocarbon receptor (AhR) enactment is reported to change the hepatic articulation of circadian clock controllers but the impact on clock-controlled metabolism has not been thoroughly investigated. This study examines the consequences of AhR activation on hepatic transcriptome and metabolome rhythmicity in male C57BL/6 mice gavaged orally with 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) every 4 days for 28 days. The effect shows that, TCDD reduces the rhythmicity level of several genes including most of the circadian clock regulated genes. One interesting observation was an increase in the amplitude of the Circadian Locomotor Output Cycles Kaput (*Clock*) gene that produces the CLOCK transcription factor that plays a crucial role in generating circadian rhythms.

1 Introduction

In response to the rotation of the Earth and the ensuing light-dark cycles, animals and plant have evolved a 24-hour (h) circadian clock that entrains physiological activities. These clocks generate cellular rhythms known as the circadian rhythms which is responsible for hormone release, digestion, body temperature, and other physical activities such as sleep-wake, feeding to specific times of the day. Also the circadian clock is intimately involved in the control of metabolic and physiological processes and its dysregulation is associated with the onset and development of numerous human diseases, including sleep disorders, depression, and dementia. The cellular regulatory network of circadian clocks involves several genes and proteins, and relies on multiple interactions, transcriptional as well as post-translational.
The core circadian clock gene network consist of the core clock activators (*Clock* and *Bmal1*), clock repressors (*Cry1*, *Cyr2*, *Per1*, *Per2* and *Per3*) and clock nuclear receptors (*Rev-Erb*$\alpha\&\beta$, *Ror* $\alpha\&\beta$, *Dbp* and *Nfil3*) . The mammalian circadian clock relies on the master genes CLOCK and BMAL1 to drive rhythmic gene expression and regulate biological functions under circadian regulation. The products of the *Clock* and *Bmal1* genes form the complex CLOCK-BMAL1 which binds to the E-box of the clock regulator genes to activate transcription. In this study, I use the JTK-CYCLE algorithm to detect and predict all Rhythmic genes in male C57BL/6 mouse liver with and the effect of TCDD on the circadian clock gene network. This will also measure the period, phase and amplitude of cycling transcripts facilitating downstream analysis. This will go a long way to extend clinical investigators knowledge of the genes that if deactivates can cause harm in the mammalian clock network, the model can also be adapted to model similar problems: This is because the efficacy of many different drugs has been shown to depend considerably on the time of the day they are administered. This discovery of these rhythmic genes leads to the construction of a molecular based dynamical model to analyze the dynamics and oscillatory patterns for these genes. Also on a broader impact A more accurate understanding of the dynamics and oscillatory mechanisms of the circadian clock and the effects of perturbation in the liver vows to furnish biomedical engineers and clinical examiners with a successful new weapons store with which to fathomthese conditions. This will lead to more efficient and simplified communication between medical practitioners and patients since the gap between medical practitioners and combating perturbations of the clock would be reduced.

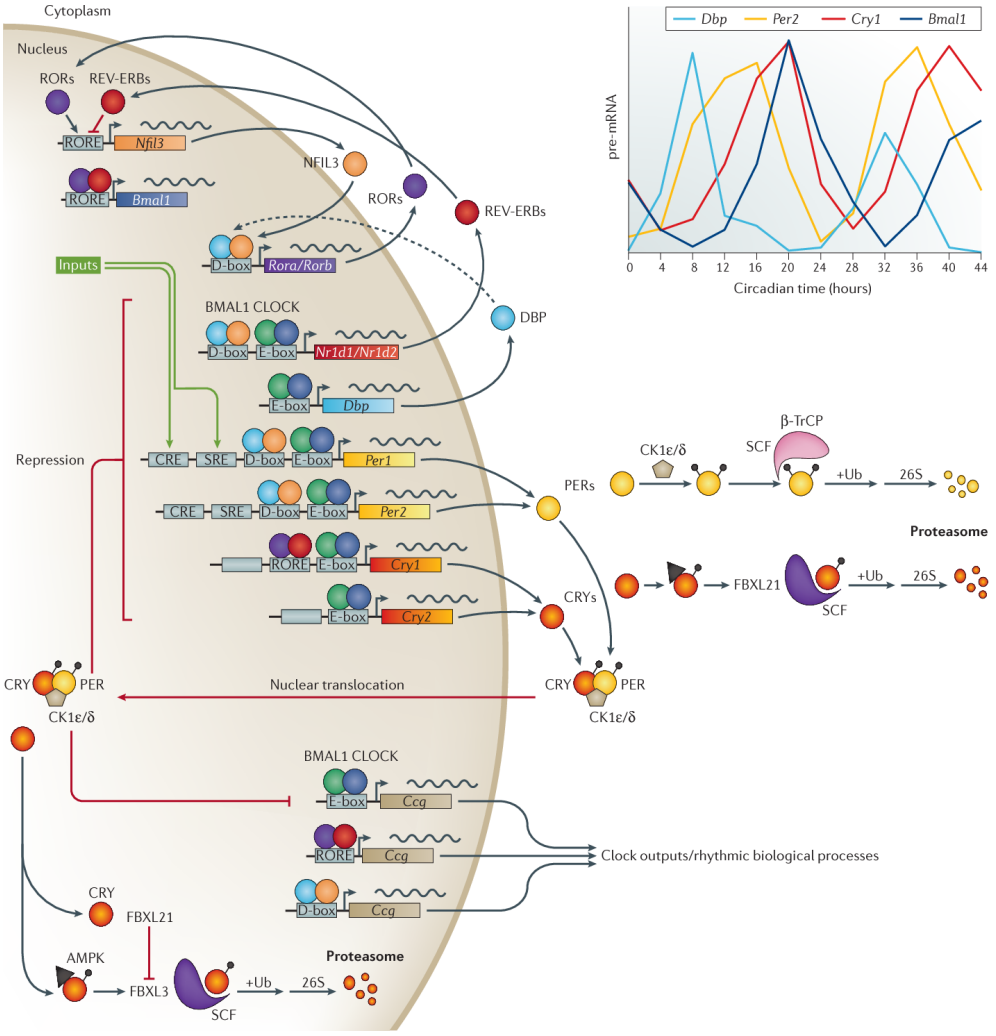


Fig 1:
Core components of the mammalian circadian clock. In the core feedback loop, the transcription factors BMAL1 (green circles) and CLOCK (blue circles) bind to E-box domains on gene promoters, including the genes for Per1 and Per2 (yellow) and Cry1 and Cry2 (red/yellow). PERs (yellow circles) and CRYs (red/yellow circles) dimerize and translocate to the nucleus after binding with CK1δ or CK1ϵ, where they repress their own transcription. The stability of PER and CRY is regulated both in the cytoplasm and within the nucleus by several proteins, including FBXL21 and FBXL3. In a second feedback loop, CLOCK and BMAL1 also regulate the transcription of genes for the nuclear receptors REV-ERBα and REV-ERBβ (red circles), which compete with the retinoic acid-related orphan receptors, RORα, RORβ, and ROR$gamma$ (purple circles) for binding to RRE elements on the BMAL1 gene promoter, providing both positive (ROR) and negative (REV-ERB) regulation of BMAL1 transcription. A third feedback loop is mediated by CLOCK/BMAL1-mediated transcription of the gene Dbp (light blue) and the ROR/REV-ERB-mediated transcription of Nfil3 (orange). DBP (light blue circles) and NFIL3 (orange circles) dimerize and bind to D-box elements on the promoters of many of the core clock genes, providing additional layers of regulation. In addition, CLOCK/BMAL1, ROR/REV-ERB, and DBP/NFIL3 regulate the transcription of many other clock output genes (Figure modified from; Takahashi 2017).

2 Data & Methods

2.1 Raw Data

Raw dataset of 48 samples consisting of RNA-Seq Analysis of Vehicle and TCDD-Elicited Changes in Circadian Hepatic Gene Expression in Male Mice were downloaded from the NCBI Gene Expression Omnibus site which was generated from Dr. Timothy Zacharewski's Lab .
The dataset consisted of 24 samples for Vehicle & 24 samples for treatment with 30 μg of 2,3,7,8-Tetrachlorodibenzo-ρ-dioxin (TCDD).

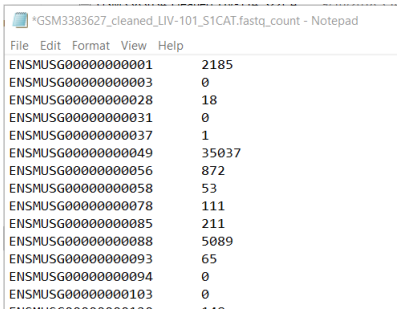


Fig 2: A sample of the Raw data

2.2 Data Proccesing and Normalization

The datasets were processed and annotated in R using The R-package BioMaRt. The normalization approached used was the Transcripts Per Kilobase Million (TPM) and trimming. After the normalization all rows withzero (0) at all columns were eliminated and those with one or two zero (0) values in their columns were infilled by using the column mean of the respective column. Using the sensitivity approach in this manner will lead to having some noise as false positives in the dataset.

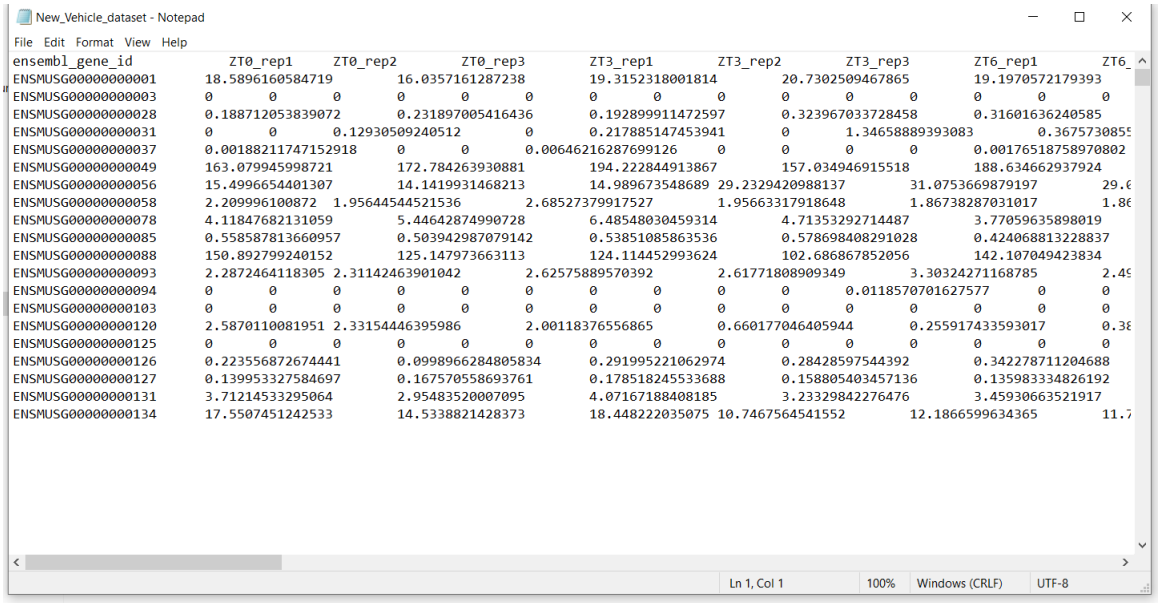


Fig 3: Vehicle Processed Data

2.3 Exploratory Analysis

Data exploratory analysis was done for both the Vehicle (Control) dataset and the treated (TCDD) dataset. This was done using the skimr package in R.

Name	Vehicle_data	character	1
Number of rows	27645	numeric	24
Number of columns	25	Missing values	None

Table 1: Vehicle Dataset Summary.

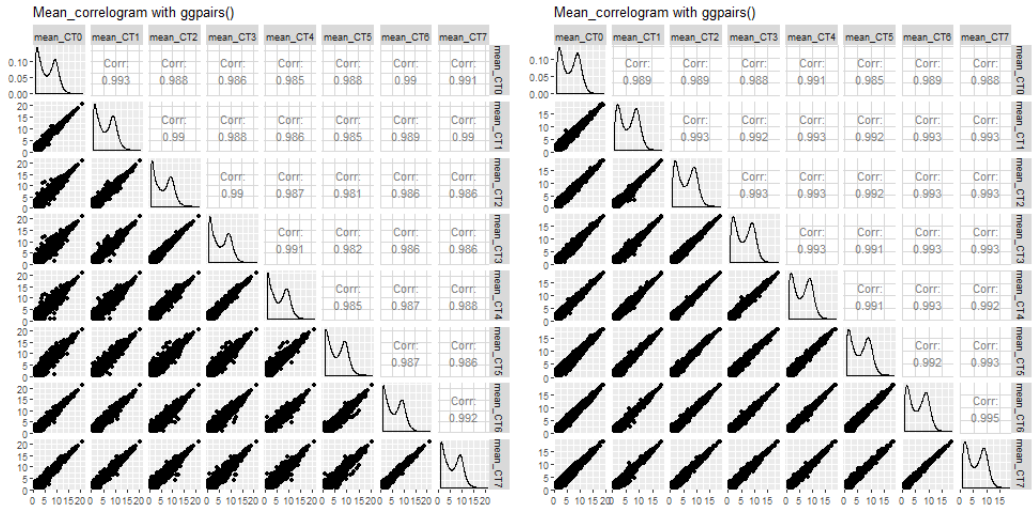
Name	TCDD_data	character	1
Number of rows	30140	numeric	24
Number of columns	25	Missing values	None

Table 2: TCDD Dataset Summary.

One interesting thing that was revealed was that the genes expressed in the treated (TCDD) dataset were more that those that were expressed in the vehicle (control) dataset.

2.4 Data Visualization

Fig 4: Correlogram of Vehicle(Control) processed dataset (Top) and Treated ()TCDD) processed dataset (bottom). Because we have eight (8) time points and three replicates for each time point, the mean of the three replicates for each time point was calculated to and plotted to represent the time point. The Correlogram presents the distribution of the datapoints in each timepoints, the correlation between the time points.



2.5 Method

Jonckheere-Terpstra-Kendall CYCLE ()JTK-CYCLE) is used to efficiently identify and characterize cycling variables in large dataset. The algorithm is also used to distinguished between rhythmic and non-rhythmic transcript more reliable and efficiently. JTK-CYCLE accurately measures the period, phase, and amplitude of cycling transcripts, facilitating downstream analyses. The JTK-CYCLE algorithm is a combination of two statistical methods:The Jonckheere-Terpstra (JT) test which is a non-parametric test that is most powerful for detecting monotonic orderings of data across ordered independent groups and Kendall's tau measure of rank correlation that is used to measure the association between two measured quantities. The Jonckheere-Terpstra-Kendall (JTK) algorithm applies the JT test to a family of alternative hypothesized group orderings, while keeping the group sizes fixed. For enhanced computational efficiency, these tests are performed by utilizing the mathematical equivalence between the exact null JT distribution and the exact null distribution of Kendall's tau correlation between a continuous random variate and an ordinal grouping factor. 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 where the group oderings are derived from the cosine curve. The optimal periods and phases found by JTK-CYCLE are highly resistant to outliers, because Kendall's tau depends only on the signs of the inter-group differences between pairs of values. JTK-CYCLE estimates the amplitude of each optimal cyclical pattern by calculating the one-cycle median sign-adjusted deviation from the median (*msad*), where each sign-adjusted deviation equals the product of the deviation and the associated sign of the optimal cosine pattern. For a perfect cosine pattern with amplitude A, the one-cycle *msad* equals the *mad* (median absolute deviation from the median) which in turn equals $\frac{A}{\sqrt{2}}$.

2.6 Software

The JTC-CYCLE R-Script code written by Micheal Hughes was used to asses the rhythmicity of the transcriptomic and metabolomic dataset. This scripts takes two inputs, an annotation file of the transcript which could be either probeset IDs, geneIDs, gene names, etc and the the second input file is the transcript at various time point. An overview of how to run the algorithm is added to the supplementary elements with the R script.

3 Results

The effect of TCDD on hepatic transcriptomic we assessed by applying the JTK-CYCLE algorithm on the RNA-Seq dataset which were generated every 3h interval over a 24h period. JTK-CYCLE analysis of the RNA-Seq data detected 4415 hepatic genes exhibiting rhythmic expression with period between the hours of 21 and 24 in the Vehicle dataset which is equivalent to 16% of the 27645 genes expressed in the liver Vehicle dataset. Also 1716 hepatic genes exhibiting rhythmic expression with period between the hours of 21 and 24 in the treated (TCDD) dataset which is equivalent to 6% of the 28716 genes expressed in the liver treated (TCDD) dataset. Specifically, 20 core hepatic clock regulators exhibited oscillating expression including (i) the Clock regulator genes that produces the clock transcription factors *Arntl* (aka *Bmal1*), *Clock*, and *Npas2*, (ii)the clock repressor genes *Per2*, *Per3*, *Cry1*, and *Cry2*, (iii) the nuclear receptor genes *Nr1d1* (encodes *REV-ERB*α), *Nr1d2* (encodes *REV-ERB*β), and *Rorc* (encodes *ROR*γ), and (iv) the D-box binding transcription factors *Dbp*, *Tef*, *Hlf*, and *Nfil3*. Hepatic expression of the RORE-binding transcription factor Rora did not exhibit rhythmicity, consistent with previous reports that it lacks circadian oscillation in peripheral tissues.

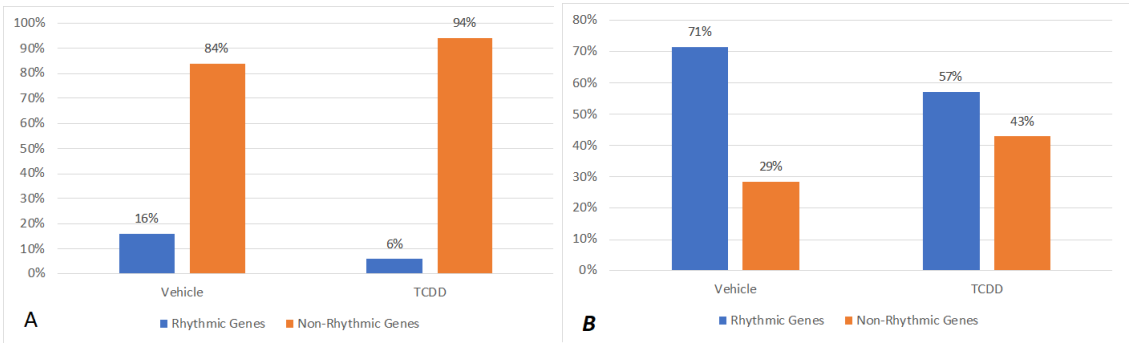


Fig 5: The effect of TCDD on the diurnal rhythmicity of hepatic gene expression in male C57BL/6 mice orally gavaged with sesame oil vehicle or 30 μg/kg TCDD every 4 days for 28 days. (A) The percentage of hepatic genes which exhibited rhythmic expression in vehicle and treated (TCDD) mice dataset. 4415 of 27645 (16%) hepatic genes which exhibited rhythmic expression in vehicle mice dataset. 27000 (94%) hepatic genes lost their oscillation pattern following treatment. (B) The percentage of clock related genes which exhibited rhythmic expression in vehicle and treated (TCDD) mice dataset. 20 of 28 (71%) clock related genes which exhibited rhythmic expression in vehicle mice dataset. 12 (43%) clock rhythmic genes lost their oscillation pattern following treatment.

Even though most clock related genes retained their rhythmicity following treatment, TCDD had an effect on these genes by dampening their rhytmicity. TCDD caused a reduction in amplitude in most of the clock gene. One of the Clock activator genes *Clock* had an increase in its amplitude after treatment. The average of the three replicates at each time point was calculated and used to illustrate the oscillation generated by each gene. TCDD increased the acrophases of *Rorc*, *Per3*, *Nr1d1* and decreased the acrophase of *Dbp*, *Nr1d2* and *Arntl*. Overall, TCDD abolished the rhythmicity of some of clock-controlled hepatic genes, while those that continued oscillating following treatment exhibited reduced amplitudes (in most cases). Beyond the collapse of the hepatic circadian clock, previous studies have shown TCDD also causes the loss of liver-specific and sexually dimorphic gene expression.

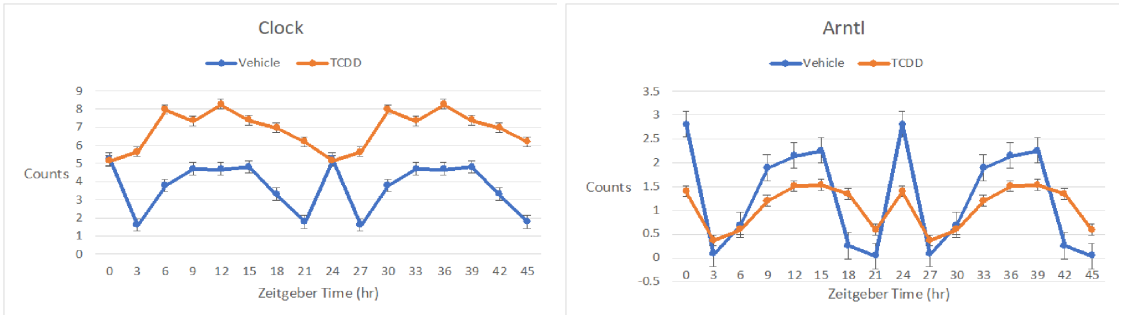


Fig 6: The effect of TCDD on the rhythmicity of core clock activator genes expression in male C57BL/6 mice orally gavaged with sesame oil vehicle or 30 μg/kg TCDD every 4 days for 28 days. TCDD increased the amplitude of the *Clock* gene but it decreased the amplitude of the *Arntl (Bmal1)* gene. Data are double plotted along the x-axis for better visualization.

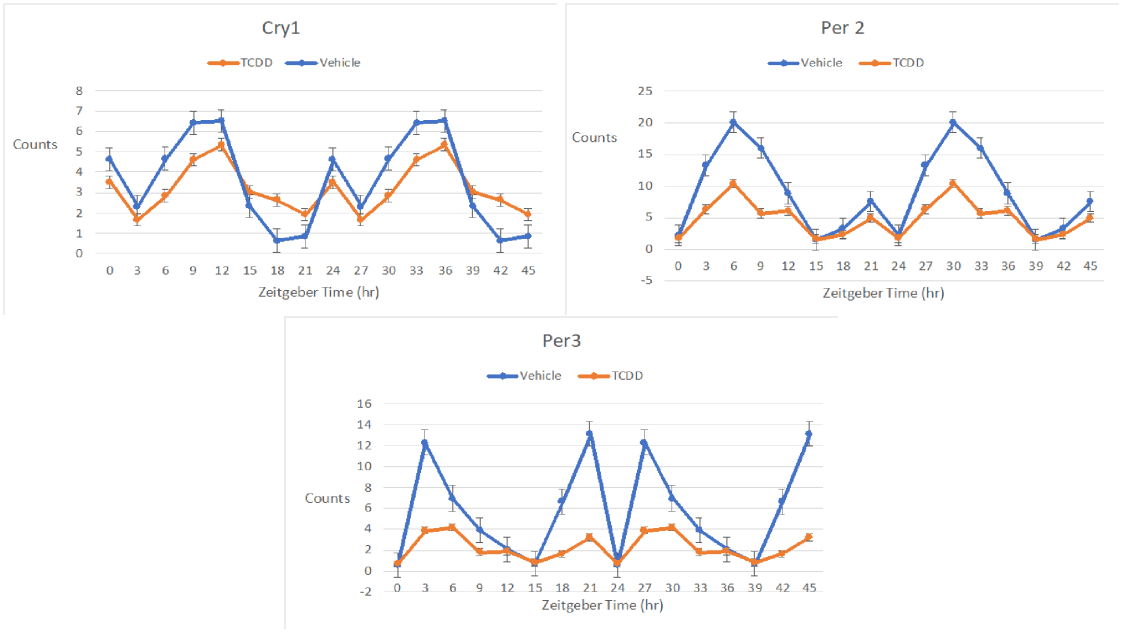


Fig 7: The effect of TCDD on the rhythmicity of core clock repressor genes expression in male C57BL/6 mice orally gavaged with sesame oil vehicle or 30 μg/kg TCDD every 4 days for 28 days. TCDD decreased the amplitude of the *Cry1* gene but no rhythmic pattern was recorded by the *Cry2* gene. Also there was a large decreased in the amplitude of the *Per2* and the *Per3* gene after treatment. Data are double plotted along the x-axis for better visualization.

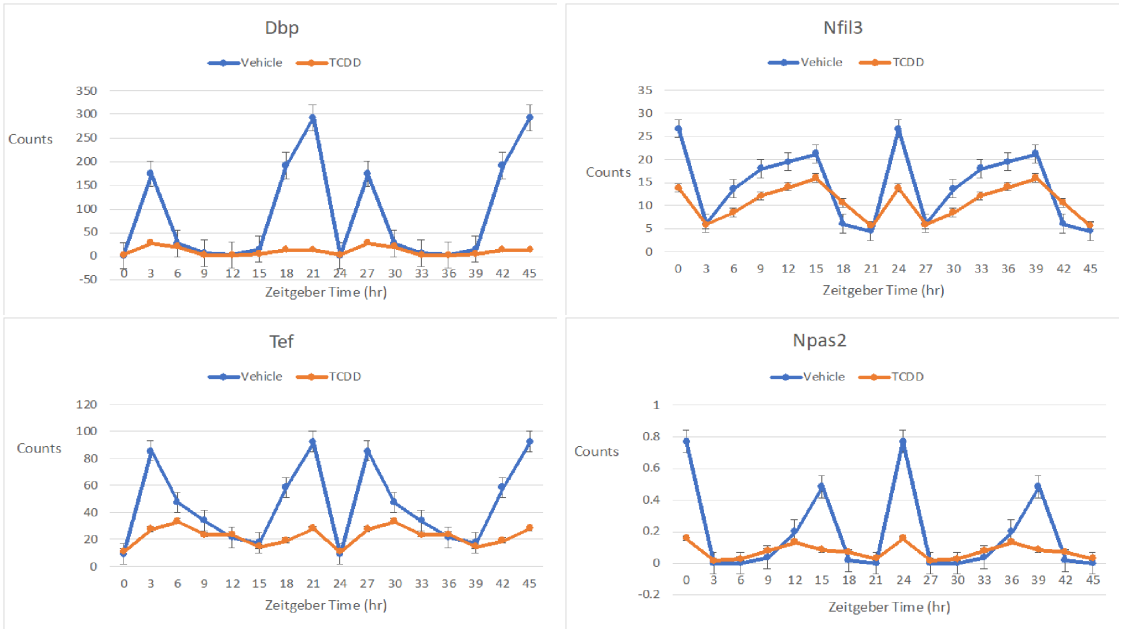


Fig 8: The effect of TCDD on the rhythmicity of the D-box binding transcription factor genes expression in male C57BL/6 mice orally gavaged with sesame oil vehicle or 30 μg/kg TCDD every 4 days for 28 days. TCDD dramatically decreased the amplitude of of all the D-box binding transcription factor genes. Data are double plotted along the x-axis for better visualization.

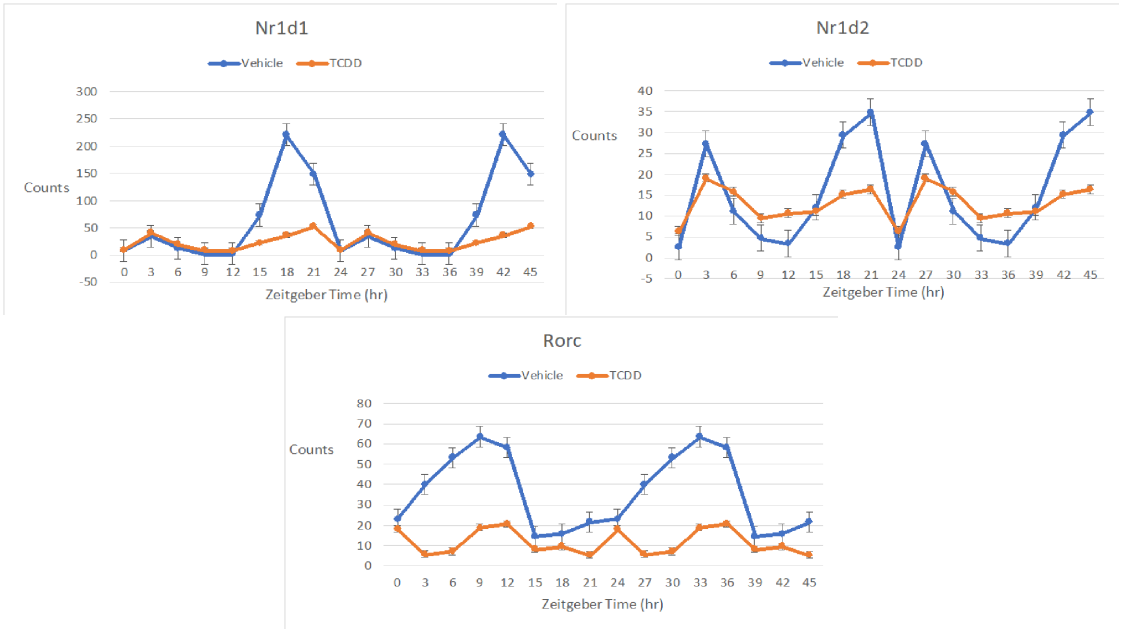


Fig 9: The effect of TCDD on the rhythmicity of the the nuclear receptor genes expression in male C57BL/6 mice orally gavaged with sesame oil vehicle or 30 μg/kg TCDD every 4 days for 28 days. TCDD dramatically decreased the amplitude of of all the nuclear receptor genes. Data are double plotted along the x-axis for better visualization.

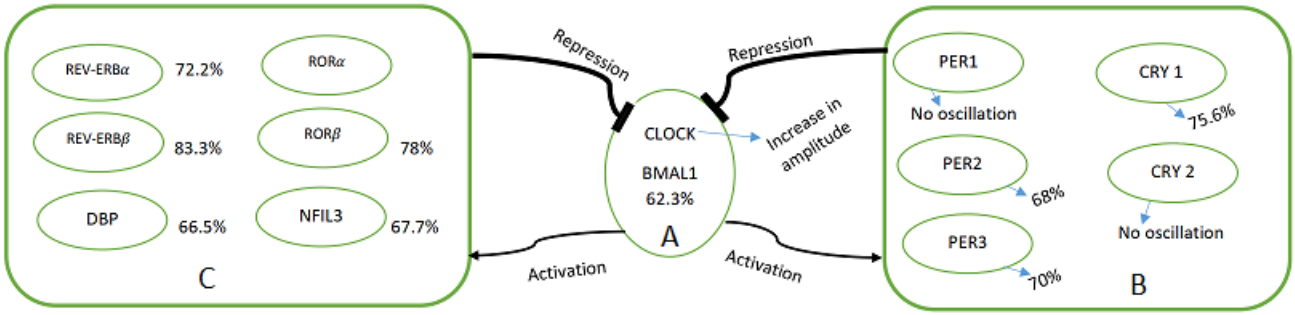


Fig 10: The percentage of amplitude reduction confirmed by JTK-CYCLE in the clock-regulated genes. This figure also shows the positive and negative feedback loop involved in the clock mechanism.

4 Discussion

The current research explored the impact of TCDD on the rhythmicity of the hepatic transcriptome and metabolome in mice specifically clock regulated genes. TCDD activation of AhR dampened the rhythmic expression of 12 hepatic core clock genes, causing either a reduction in amplitude or total oscillation loss. TCDD cause changes in the amplitude, phase and period of the hepatic transcriptome and metabolome in mice. In addition to reduced transcriptional rhythmicity of the core clock regulators, TCDD altered cues that entrain the hepatic clock. The

hepatic peripheral clock can be reset by specific dietary nutrients such as glucose and amino acids, allowing synchronization with feeding times.

5 Future Directions and Limitations

This project was a start to my bioinformatics approach of modeling the circadian clock. A decrease in the amplitude in *Arntl* caused a decrease in amplitude in most of the clock regulatory genes. What to do next is to get a *Bmal1* Chip-Seq bed file, extract the transcription start and end points for the clock regulatory genes, extract their nucleotides within these ranges by using the mouse genome fasta file. After getting this file then check for DRE-AHR-peaks and AHR binding sites within the promoter region of these genes. This will help investigate the question of TCDD preventing *Bmal1* to bind to the promoter regions of the clock genes that decreased in amplitude by sitting on the AHR binding sites in the promoter region of these genes.

I had a lot of challenge plotting the rhythmicity of the genes in R script so I used excel which was first time using excel for plotting. This helped me learn another approach to plot graphs.
I still lack much of the knowledge in understanding the circadian clock and I am still challenging myself to learn more about the clock regulatory genes and their interactions.

6 References

[1] Ukai, H., & Ueda, H. R. (2010). Systems biology of mammalian circadian clocks. Annual review of physiology, 72, 579-603.
[2] Hughes, M. E., Hogenesch, J. B., & Kornacker, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. Journal of biological rhythms, 25(5), 372-380.
[3] Takahashi, J. S. (2015). Molecular components of the circadian clock in mammals. Diabetes, Obesity and Metabolism, 17, 6-11.
[4] Hayes, K. R., Baggs, J. E., & Hogenesch, J. B. (2005). Circadian clocks are seeing the systems biology light. Genome biology, 6(5), 219.
[5] Wichert, S., Fokianos, K., & Strimmer, K. (2004). Identifying periodically expressed transcripts in microarray time series data. Bioinformatics, 20(1), 5-20.
[6] Levine, J. D., Funes, P., Dowse, H. B., & Hall, J. C. (2002). Signal analysis of behavioral and molecular cycles. BMC neuroscience, 3(1),1.
[7] Ren, Y., Hong, C. I., Lim, S., & Song, S. (2016). Finding clocks in genes: a Bayesian approach to estimate periodicity. BioMed research international, 2016.