ENCORE: A Visualization Tool for Insight into Circadian Omics

Hannah De los Santos
Institute for Data Exploration and
Applications/Department of
Computer Science
Rensselaer Polytechnic Institute
Troy, NY
delosh@rpi.edu

Kristin P. Bennett
Institute for Data Exploration and
Applications/Department of
Computer Science
Rensselaer Polytechnic Institute
Troy, NY
bennek@rpi.edu

Jennifer M. Hurley
Department of Biological
Sciences/Center for Biotechnology
and Interdisciplinary Sciences
Rensselaer Polytechnic Institute
Troy, NY
hurlej2@rpi.edu

ABSTRACT

Circadian rhythms are 24-hour biological cycles that control daily molecular rhythms in many organisms. The cellular elements that fall under the regulation of the clock are often studied through the use of omics-scale data sets gathered over time to determine how circadian regulation impacts cellular physiology. Previously, we created the ECHO (Extended Circadian Harmonic Oscillator) tool to identify rhythms in these data sets. Using ECHO, we found that circadian oscillations widely undergo a change in amplitude over time and that these amplitude changes have a biological function in the cell. However, ECHO does not align gene ontologies with the identified oscillating genes to give functional context. Thus, we created ENCORE (ECHO Native Circadian Ontological Rhythmicity Explorer), a novel visualization tool which combines the disparate databases of Gene Ontologies, protein-protein interactions, and auxiliary information to uncover the meaning of circadianly-regulated genes. This freely-available tool performs automatic enrichment and creates publication-worthy visualizations which we used to extend previously-gathered data on circadian regulation of physiology from published omics-scale studies in three circadian model organisms: mouse, fruit fly, and *Neurospora crassa*.

CCS CONCEPTS

• Applied computing → Bioinformatics;

KEYWORDS

circadian rhythms, proteomics, transcriptomics, gene expression, gene enrichment, protein-protein interaction networks, visualizations

ACM Reference format:

Hannah De los Santos, Kristin P. Bennett, and Jennifer M. Hurley. 2019. ENCORE: A Visualization Tool for Insight into Circadian Omics. In *Proceedings of 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Niagara Falls, NY, USA, September 7–10, 2019 (ACM-BCB '19)*, 10 pages.
<https://doi.org/10.1145/3307339.3342137>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6666-3/19/09...\$15.00

<https://doi.org/10.1145/3307339.3342137>

1 INTRODUCTION

Fundamental to the comprehension of biological function is an understanding of circadian rhythms, which are timed by a highly-conserved molecular circuit, or clock. This "clock" dictates a significant portion of organismal physiology [21, 22]. In eukaryotic systems, the clock is composed of a transcription/translation negative feedback loop (TTFL) with a roughly 24-hour period [27]. This TTFL imparts oscillations on many physiological parameters that affect a broad range of organismal functions, from the regulation of metabolism to physical/cognitive abilities, allowing an organism to be more competitive in the earth's day/night cycle [23, 32, 33]. Due to the large number of cellular systems under the control of the clock, disruptions of our circadian rhythms can profoundly influence health and well-being, with chronic disruption of the clock enhancing the risk of many diseases, including cancer, diabetes, heart disease, and stroke [24]. The import and conservation of molecular rhythms has driven researchers to identify which cellular components fall under circadian regulation to understand the broad impact of the circadian clock on biology.

To identify clock-regulated components, samples gathered over circadian time are used to enumerate the levels of many of the molecular constituents of the cell over the circadian day, most commonly mRNA (transcriptome) and protein (proteome). From this data, circadian rhythms are tracked using the oscillatory waves they resemble. While in the past, fixed-amplitude computational tools such as JTK_CYCLE [31] and eJTK_CYCLE [34] were used to categorize circadian genes, these methods largely ignored the prevalence of changing amplitudes in these data sets [4, 58]. To compensate for this, we designed the Extended Circadian Harmonic Oscillator (ECHO) program to take this amplitude change (AC) into account to robustly detect circadian genes [17, 18]. ECHO demonstrated that each of the AC categories (damped, harmonic, and forced) has specific biological functions, highlighting a new level of biological complexity in circadian regulation [17, 18]. However, while ECHO improved the identification of oscillating elements, it also demonstrated that a large portion of the transcriptome and proteome were found to oscillate, anywhere from 5–80%, making the functional impact of those oscillating elements difficult to discern from the considerable amount of data that is gathered.

To ascertain the functional impact of these oscillations, researchers often employ two repositories: the Gene Ontology database (GO) [3] and the Search Tool for the Retrieval of Interacting Genes/Protein database (STRING) [46]. GO, founded by the Gene Ontology Consortium, was created in the wake of wide genomic sequence availability [3]. GO is structured as a directed acyclic graph, a hierarchy

in which terms at higher levels indicate more broad function and those at lower levels are more specific. GO utilizes an ontological structure and clearly-defined vocabulary to allow the unification of conserved functional vocabulary between organisms, with each gene potentially contributing to many different functional terms. GO is the source for the derivation of significant functions of a specific group of genes through GSEA and many websites will calculate enrichments automatically for a variety of organisms [28, 44, 50, 52]. STRING is a repository of protein-protein interactions for many different model organisms, with the idea that an understanding of which proteins interact with one another will lead to a clearer picture of biological function [46]. Interaction data comes from several different sources and is scored based on confidence in the interaction, with a higher number indicating more confidence [25].

While existing tools based on GO and STRING are valuable for deriving an understanding of the impact of circadian regulation, these tools present several limitations. First, to our knowledge, these tools have not been integrated for circadian biology, which could enhance their impact. Second, the tools that identify GO enrichment provide overwhelming lists of terms, with little to no visualizations to clarify output. It is well known that visualizations are key for humans to understand data; when made with humans in mind, visualizations strongly enhance the amount of understanding derived, as well as the insight towards future experimentation and analysis [53]. While websites such as REVIGO and WEGO provide reduction and visualizations of GO terms, these do not provide deep dataset exploration, as they only summarize GO results [51, 59]. Finally, the identification of the biological relevance of AC categories suggests a need to include these in GO and STRING analysis, not reflected in previously mentioned tools [17, 18, 47, 49].

We therefore created the ECHO Native Circadian Ontological Rhythmicity Explorer (ENCORE), an application which combines the power of GO and STRING with associated statistical-enrichment testing to derive biological understanding of circadian function. By leveraging AC categories provided by ECHO, we are able to find distinct functional AC differences for a variety of organisms, utilizing a user interface that allows more thorough exploration of enrichments by integrating protein-protein interactions. ENCORE also connects to outside databases, UniProt [54] and QuickGO [7], to provide access to further information about genes and GO terms. In total, we find that the application of ENCORE to large-scale circadian data sets allows for a deeper understanding of the impact of circadian regulation over biological functions.

2 METHODS

We created ENCORE, a methodology combining several disparate databases in order to derive further insight from circadian rhythms through novel visualizations, available on Github¹.

2.1 Preprocessing and Gene Set Enrichment Calculation

After circadian oscillations were detected by ECHO [18] (Figure 1), our methodology began by entering the ECHO results into the ENCORE interface to automatically generate ontological gene-set enrichments for all genes that oscillate with a circadian period, as

¹<https://github.com/delosh653/ENCORE>

Figure 1: ENCORE integrates several databases to visualize circadian omics data. After ECHO detects rhythmic genes, ENCORE combines information from GO, STRING, and auxiliary information from QuickGO and UniProt to derive 2D and 3D visualizations of ECHO analyzed omics-scale data sets. Data shown from Hughes et al. [30] and Li et al. [40]

well as subsets of this group divided by amplitude change (AC) category. In order to calculate enrichments relative to a wide background, as by the traditional GSEA method [50], genes represented in the whole genome for that organism and unrepresented within the user's data were appended, with a designation of not circadian. Gene names were then mapped to the designated preferred ID by ontology package. Missing or duplicated names were then removed. STRING names were mapped based on preferred ID, as computed by their R package *stringdb* [25]. We have enabled ENCORE to work with many common model organisms with R packages, including human [11], mouse [12], *N. crassa* [42], fruit fly [9], mosquito [8], *E. coli* [10], and *S. cerevisiae* [13].

For each AC category, as well as the total set of circadian genes, gene ontological enrichments were automatically calculated using topGO [1]. First, circadian genes were filtered by each category through a user specified p-value cutoff, p-value adjustment scheme (such as Benjamini-Hochberg), and period restriction. After annotations to given genes were matched through topGO, the classic Fisher's statistic was calculated for all terms and multiple hypothesis adjusted for all terms with at least two attributed significant genes. The resulting gene ontological hierarchy was then pruned to contain only significant GO terms or insignificant parents leading to significant GO terms. Other pertinent information, such as fold enrichment, was also calculated, then exported to the user as an .RData file for ontological exploration through the ENCORE interface. This file, which contains all necessary enrichment and protein-protein interaction annotations, can also be examined separately with R if desired.

Figure 2: ENCORE generates interactive visualizations to understand the impact of circadian regulation over cellular biology. A. Ontology Map: a Sankey diagram showing the damped GO term hierarchy for the biological process "cellular macromolecule metabolic process". Hovering over each bar reveals the full title of that ontological term. Hovering over each link shows the parent and child GO-term relationship. **B. Ontology Explorer:** a stacked bar graph of AC categories contributing to the significant child terms of "cellular macromolecule metabolic process", with the ontological parent-child path detailed at the top right. Hovering over each bar layer reveals the fraction annotated by that AC category, p-value, and fold enrichment. Hovering over each GO term bar reveals its full name and the total fraction annotated by the selected categories. **C. Group Comparison Tool:** a chord diagram and heatmap for the GO term "cellular macromolecule metabolic process". Hovering on the outermost AC-colored arc reveals only protein-protein connections related to one AC category. Hovering over each gene arc reveals protein-protein connections and expression heatmaps for only that gene, as well as its name and the total number of connections. Hovering over each chord indicates the name and AC category of the connection, as well as the heatmap of expression for the connected gene. Hovering over a genes heatmap reveals its peak phase, in hours, called "hours shifted". Data shown from Hughes et al. [30]

2.2 The ENCORE Application

Once ontologies are calculated, the user can discover gene function and interactions through several different visualizations that bring together knowledge from various databases (Figure 1); these include Ontology Maps, the Ontology Explorer, and the Group Comparison Tool. These interactive visualizations and interface are implemented in the freely available R with r2d3 [41] and shiny [14], respectively.

2.2.1 Ontology Maps. For each AC category, selecting an ontological term visualizes all paths to that term through a Sankey diagram, flowing from top to bottom (Figure 2A). Possible paths from the ontological type to the selected term also flow from top to bottom, subsetting the large hierarchical structure to an easily digestible graph. Not significant terms appear as grey, while significant terms are colored in cool colors by their AC group. In order to find more information on a specific ontological category's interactions and child hierarchy, users may click the bar near the

ontological term; this will update the "Ontology Explorer" and "Group Comparison" visualizations accordingly, allowing users to dig further into the GO term's children and connections, deriving these visualizations from the shortest ontological path between the selected child and its root ontology.

Hovering and clicking provides additional information about the ontological term (Figure 2A). To see the full length title, users may hover on the bar near the corresponding GO term to see its title. Users may hover on a link between bars to see the source ontological term of the target ontological term. Clicking on the GO term provides more information about that term through QuickGO via a link which appears in the "Gene/Term Explorer" tab of the ENCORE application [7].

2.2.2 Ontology Explorer. In order to explore each ontological term more fully and compare relative enrichments between terms, users may step through their ontology using the Ontology Explorer

(Figure 2B). After selecting AC categories, or all circadian genes, the user may start at either the root GO term, such as biological process, or jump to a specific point in the desired ontology through the Ontology Map. Upon selecting the GO term, the Ontology Explorer appears as a layered bar graph representing the fraction annotated of chosen AC categories for the selected GO terms's children of the total amount of annotated genes. The width of each column represents the total fraction annotated relative to other child terms, sorted by relative abundance. To delve further through the hierarchy, users may click on the corresponding warm-colored node below each column to see that term's children, allowing for an in-depth exploration of the ontological hierarchy.

The path taken through the ontology hierarchy, as well as the original data set, appears in the top right corner, with the current parent category selected at the highest place in bold (Figure 2B). To step back, users click on the corresponding "go back" arrow to the left of the layered bar chart. To switch between viewing significant and nonsignificant child categories, users click on the star on the left, in order to click to possible significant child terms.

Exact numerical enrichment information appears upon hover and click interactions (Figure 2B). Users can hover on each AC category's bar layer to see the fraction annotated, p-value of enrichment, and fold enrichment for each category. To get the total fraction annotated by the representative groups, users hover over each ontology term bar. As with the Ontology Map, users may click the GO Term for more information about that term as provided by QuickGO.

2.2.3 Group Comparison. Once a GO term is selected through either the Ontology Map or Ontology Explorer, the user can explore STRING-determined protein-protein interactions of genes enriched for that term in the Group Comparison Tool (Figure 2C). These protein-protein interactions are represented by a chord diagram, where each chord represents the presence of an interaction between two genes. Chord colors that are darker than their AC group color indicate connections between genes within the same AC category, while chords that are connected between AC groups maintain their category color. Chords are selected by highest STRING score, and users can find more connections, if any, by increasing the maximum number of protein-protein interactions. Furthermore, mean-centered, normalized heat maps of gene expression sorted by phase within each AC group appear on the outside of the chord diagram with the first time point towards the center. This visualizes the peak phases of each gene in the protein-protein interaction network, creating a circadian "clock" of interactions.

Hovering and clicking on heatmap chords reveals more information about the genes involved in the protein-protein interaction (Figure 2C). Hovering on the outermost colored arc displays all the chords in only one AC coefficient category, highlighting the density of chord connections between that AC group and other categories. Further hovering on any gene-specific arc restricts all chords and heatmaps to those associated with the specific gene, with the gene name and number of connections appearing above the arc. Users can then hover on the chord to display the connection to that gene through their heatmap. Hovering on the heat maps themselves displays the phase shift, in hours, for the corresponding gene. Clicking on a gene arc or a chord displays information

about either that gene or its connecting gene, respectively, in the Gene/Term Explorer tab (see Auxiliary Information), including a link to external information on UniProt [54].

2.2.4 Auxiliary Information. In addition to the main exploratory visualizations, various pieces of auxiliary information are available through the Gene/Term Explorer and Data Information tabs. When users click on an ontological term in the Ontology Map/Ontology Explorer or a gene in the Group Comparison, links to information about the term appear in the Gene/Term Explorer, sourced from QuickGO and UniProt respectively. Further, clicking on a gene through the the Group Comparison Tool adds additional gene information. In this tool, the temporal change in levels of a given gene, including ECHO fit and a summary of the ECHO parameter information, appears for the selected gene. Genes appearing in the chord diagram are displayed, along with their AC categories and hours-shifted (phase) values. Information about the data set and ENCORE enrichment settings also appears in a separate "Data Information" tab.

2.3 ENCORE as a 3D Interface

We have also created a 3D interface for this work through the use of the Rensselaer Campfire [2], a novel multi-user, collaborative, immersive computing interface that is ideal for group interaction with data visualizations and explorations. The Campfire is a 3D cylindrical structure, consisting of two separate windows: a circular "floor" and a cylindrical "wall". This campfire is also connected to two adjoining monitors that provide additional information, as well as an external tablet computer to act as a "controller". This structure allows users to more accurately compare and visualize their results, as 2D-warping can happen due to converting from radial to Cartesian coordinates. Further, the Campfire's structure inherently allows collaboration and discussion, enhancing the discovery of novel ontological insights.

In the 3D Campfire version of ENCORE, there are several options to display the data (Figure 3A). Initially, the Ontology Map is placed on the user's controller, along with other user inputs, to start the ontological exploration for the given data set. From there, users can choose to switch between the Ontology Navigator and the Group Comparison Tool in the Campfire itself. In the Ontology Explorer mode, a pie chart representing the total fraction of genes attributed to the GO term divided into the significant child categories will be displayed on the floor. The outer wall will display the stacked bar graph of the fraction annotated for each AC category (Figure 3B). Further information is also available through hover and click interactions. Users can click on one of the pie slices to move forward through the ontology, and a center arrow to go back.

After a GO term is selected by clicking through either the Ontology Map or a pie slice through the Campfire, the user can choose to switch to the Group Comparison Tool (Figure 3C). In this case, the protein-protein interaction chord diagram is displayed on the floor while the expression heatmap is displayed on the wall. Hover and click interactions are retained from the 2D version, including gene chord emphasis and numerical information.

The external campfire monitors display the visualization not currently in the floor of the campfire, i.e. while the Ontology Explorer is in the campfire, the Group Comparison Tool is displayed on the

first external monitor and vice versa (Figure 3A, D). The second external monitor displays all auxiliary information from the campfire interactions, including the gene expression information, chord diagram genes, and external QuickGO and UniProt information (Figure 3D).

2.3.1 Connecting Campfire Windows with mwshiny. As the Campfire is essentially multiple windows (floor, wall, monitors, and controller), we needed to create a user interface that could span these windows interactively – if one input changes in one window, the rest should update accordingly. While Shiny provides our necessary level of interaction, as seen in the 2D version, it does not extend over more than one window. Thus, we have created an R package called *mwshiny*, which extends the Shiny paradigm to include multiple windows, called Multi-Window Shiny [19]. This package extends the functionality of Shiny across multiple disparate windows utilizing only Shiny's functions, creating a lightweight but powerful package. By elegantly utilizing Shiny's reactive structure, we break down app development into a simple workflow with three parts: user interface development, server computation, and server output. This makes our use ideal for the ENCORE Campfire, as *mwshiny* spans our five windows easily.

2.4 Data

To demonstrate ENCORE's efficacy, we chose to confirm and extend the results of three previously published papers across several model systems in circadian biology: Solenas et al. (2017) [49], Li et al. (2019) [40], and Hurley, et al. (2018) [33].

2.4.1 Solenas et al. (2017). Solenas et al. (2017) [49] demonstrated differences in cycling transcripts between adult and aged (greater than 18 months) *Mus musculus* (mouse) stem cells by taking epidermal stem cells samples every 4 hours over 24 hours in light-dark conditions with 4 replicates via microarrays. The Jonckheere-Terpstra-Kendall algorithm (JTK_CYCLE) [31] was used to determine rhythmic transcripts in the dataset, with a permutation-based p-value of 0.05 as the cutoff. Solenas et al. performed gene ontology analysis through Genomatix, using the Gene Ranker package, specifying only biological processes and an unadjusted p-value of 0.01 was used as the significance threshold.

To process the microarray dataset, we downloaded the original CEL files through GEO at GSE84580. An appropriate RMA normalization was run on the CEL files using Expression Console software from Affymetrix, then downloaded and mapped to Gene Symbols. Probes mapping to the same Gene Symbol were then averaged. The ECHO application was used with the optional preprocessing steps of smoothing, removing unexpressed genes with zero expression for 70% of the time course [29], and z-score normalization. The period was specified as 24. For this study, we also utilized a BH-adjusted p-value of 0.05 to determine rhythmicity before the resultant data was run through ENCORE with a p-value cutoff of 0.01.

2.4.2 Li et al. (2019). Li et al. (2019) [40] examined the difference in the *Drosophila melanogaster* rhythmic transcriptome in males and females in an *Achl* knockdown, as well as a wildtype, background. They harvested RNA samples from fly heads every 2 hours for 48 hours with a single replicate in constant darkness, analyzing the data with RNAseq. Li et al. used JTK_CYCLE [31] to determine

Figure 3: Using ENCORE in the Rensselaer Campfire provides a collaborative and immersive method for exploring the impact of the clock on organismal physiology. A. While the Ontology Explorer appears in the Campfire, the Group Comparison Tool appears on the external monitor. B. The Ontology Explorer is represented by a circular pie chart of component child terms on the Campfire floor, with the stacked bar graph representing the annotated fractions on the wall. C. The Group Comparison Tool has its chord diagram on the Campfire floor and its expression heat maps on the wall. D. External monitors display the visualization not appearing in the campfire, as well as auxiliary information.

rhythmicity, with an unadjusted p-value cutoff of 0.001. Ontology analysis was performed with DAVID [28], and an unadjusted p-value cutoff of 0.05 was used to determine significant enrichment.

Downloading the data from GEO at GSE120100, we used gene symbols as indicators and then analyzed the data with ECHO. The period was specified as 24, and data was smoothed and normalized. Genes with less than 70% nonzero expression prior to preprocessing were removed [29]. Rhythmic genes were identified with an unadjusted p-value cutoff of 0.001 before the resultant data was run through ENCORE with a p-value cutoff of 0.05.

2.4.3 Hurley et al. (2018). Hurley et al. (2018) identified significant differences in the rhythmic transcriptome and proteome in *Neurospora crassa*. *Neurospora* was sampled in constant dark conditions every 2 hours over 48 hours with 3 replicates using RNAseq and TMT-MS. eJTK_CYCLE (eJTK) [34] was used to determine rhythmicity with an unadjusted p-value of 0.05, as eJTK's implementation of p-value adjustment was found to be too conservative.

NCU numbers were used as gene names. As in the original study, we utilized LIMBR to remove batch effects and impute missing values [16]. This processed data was then run through ECHO, with preprocessing to remove unexpressed genes, and smooth/normalize the data. Periods were specified to be between 20 and 24 hours, as the true period for *Neurospora* is known to be 22.5 [48]. We utilized a BH-adjusted p-value cutoff 0.05, as our implementation (using p.adjust in R [45]) is not conservative.

We also reference analysis of the *Neurospora* transcriptome performed in preliminary ECHO work by De los Santos, et al. (2017) [17]. In both this work and the original analysis, functional category analysis via FungiFun [44] rather than GO was used. However, the general functional concepts are analogous to those in GO. Due to this difference, we utilize a BH-adjusted p-value cutoff of 0.05 rather than the unadjusted p-value of 0.05 used in the original studies.

2.4.4 Data Processing Considerations. It should be noted that, while each of these papers did not use p-value adjustments, it is strongly recommended to do so in all situations, including rhythmicity determination and enrichment analysis. However, we follow the conventions proposed in each of the papers in order to demonstrate ENCORE's ability to replicate and enhance previous results.

3 RESULTS

3.1 Solenas et al. (2017)

In Solenas et al., the authors found that clock-regulated genes changed in aged mice [49]. For genes rhythmic strictly in adult mice, enriched GO terms corresponded to homeostatic functions, whereas genes rhythmic strictly in aged mice corresponded to stress response genes. Genes common to both were enriched in circadian processes, DNA replication, and mitosis. We sought to replicate and extend these results by exploiting ENCORE's novel approach to connecting protein-protein interaction knowledge to GO terms.

ECHO identified 4,898 circadian genes in the adult mouse and 3,550 circadian genes in the aged mouse. While ECHO categorized genes into different AC categories, we did not analyze these categories as the time course length did not allow for amplitude analysis. Our comparison between groups found a large amount of uniquely identified genes (adult mice: 3,927 genes, aged mice: 2,579) with 971 genes in common. As expected from previous work [17], ECHO identified more circadian genes than JTK. However, the proportions between aged and adult mice remain similar, as in the Solenas et al. paper they found 2395 unique adult genes, 1560 unique aged genes, and 749 genes in common.

Like in Solenas et al., ENCORE identified many enrichments in genes rhythmic in adult, aged, or both, mice. However, ENCORE extended the depth of this data. For example, in adult mice, ENCORE validated enrichment of the "response to oxygen levels" term

Figure 4: ENCORE's chord diagrams provide further insight into the aging circadian clock's regulome. Overall chord diagrams, gene specific chord diagrams, and mRNA levels of highlighted genes in the ontological categories of A. response to oxygen levels in adult mice and B. DNA repair in aged mice. C. Overall chord diagram of the ontological category of circadian regulation of gene expression in both aged and adult mice along with a table of these genes and their number of interactions. Data from [49].

(23.33% annotation, enrichment p-value=3.69e-3, and fold enrichment 1.454) (Figure 4A) but went further with the chord diagram to highlight two genes with large relative amounts of protein-protein interactions in this process: *Myc* and *Akt1*. In the aged-mice only category, ENCORE verified the finding of enrichment in DNA damage and DNA repair (14.42% annotation, enrichment p-value=2.89e-3, and fold enrichment 1.454) (Figure 4B). In the chord diagram, 3 highly-connected genes were identified: *Mre11a*, *Rpa1*, and *Brca1*. For genes rhythmic in both sets, we repeated the findings that enriched ontologies included circadian rhythms, DNA replication, and mitosis. For circadian regulation of gene expression (22.03% annotation, enrichment p-value=6e-7, and fold enrichment 5.3719), the ENCORE chord diagram demonstrated relatively equal amounts of protein-protein interactions for the major players for the central mouse clock (e.g. *Cry1* and 2, *Per1-3*, and *Arntl*) [43] (Figure 4C).

Figure 5: ENCORE’s inclusion of AC categories highlights differences in male/female *Drosophila* circadian regulation. A. Bar plots, chord diagrams, and mRNA levels of highlighted genes in: A. wing disc development, B. male courtship behavior, and C. oxidation reduction process (male only C1. and female only C2.). Data from [40].

3.2 Li et al. (2019)

Li et al. found significant differences in rhythmic transcripts between the two sexes of *Drosophila* in wildtype flies [40]. The genes that oscillated only in male flies were related to development and sexually dimorphic behaviors. While there was little reported enrichment in female flies, oxidation/reduction processes were enriched in both males and females. However, the genes oscillating in this pathway were distinct.

We reanalyzed this data set with both ECHO and ENCORE using only the male and female control flies. ECHO identified 815 circadian genes (541 damped, 265 forced, and 548 harmonic) in the female flies and 1,812 circadian genes (885 damped, 382 forced, and 545 harmonic) in the male flies. Our comparison between groups again found a large amount of uniquely identified genes (female flies: 943 genes, male flies: 1,401 genes) with 411 genes in common. This was dissimilar to Li et al., which found proportionally less genes in common, though our findings replicated the male domination in uniqueness; 36 genes were found uniquely in females, 85 uniquely in males, and 90 genes were rhythmic in both. This is

likely due to finding many more rhythmic genes which seems to be due to high levels of forcing and damping in this data set.

ENCORE found enrichment in the same ontological pathways as Li et al. [40] but was able to extend the original findings by delineating between damped, harmonic, and forced genes. For example, Li et al. reported an enrichment in transcript oscillations in only male flies in wing disc development [40]. ENCORE demonstrated that this category is comprised of exclusively damped or forced genes (damped: 7.26% annotation, enrichment p-value=1.494e-2, and fold enrichment 1.5579, forced: 3.23% annotation, enrichment p-value=2.964e-2, and fold enrichment 1.8518) (Figure 5A) with two highly-connected genes reported: *shg* and *hh*. Conversely, genes in courtship behavior in male flies only are exclusively harmonic (6.19% annotation, enrichment p-value=1.484e-2, and fold enrichment 2.7131) (Figure 5B) and the chord diagram demonstrates that two genes are connected with over half of all other genes: *fru* and *Gr2a*. In the oxidation/reduction processes, genes rhythmic in males or females only are enriched (males only exclusively damped: 6.62% annotation, enrichment p-value=1.6629e-2, and fold enrichment 1.4212, females only damped: 4.24% annotation, enrichment

p-value=3.56e-3, and fold enrichment 1.7869, females only harmonic: 3.74% annotation, enrichment p-value=1.471e-2, and fold enrichment 1.6479) but different genes are involved in each pathway (Figure 5C1 and 2). This is evident from the chord diagrams, as *CG5599* and *kdn* are highly connected in male only pathways while *CG9314* is highly connected in female only pathways (Figure 5C1 and 2).

3.3 Hurley et al. (2018)

Hurley et al. demonstrated a broad difference between the circadianly-controlled transcriptome and proteome, including the enriched ontologies [33]. Further, in combination with De los Santos et al. (2017) [17] and Hurley et al. 2014 [32], differences in enriched ontologies were noted between AC categories. Functional categories relating to metabolic processes were enriched in the force subset, whereas functional categories relating to protein synthesis were enriched in the damped subset.

Here too, we reanalyzed the data with ECHO and ENCORE. We found 6,685 rhythmic transcripts (3,428 damped, 1,069 forced, and 2,188 harmonic) and 2,146 rhythmic proteins (935 damped, 594 forced, and 617 harmonic). The ratio of roughly three times more rhythmic transcripts to rhythmic proteins reflected to the original study, where 3,858 transcripts and 1,273 proteins were oscillatory.

As in Li et al., we found a large percentage of damped genes for the circadian transcriptome. Of the categories enriched for this damped subset, we confirmed the enrichment of several GO terms relating to protein synthesis, e.g. translation and ribosome biogenesis. Mirroring this, the translation GO term is comprised of exclusively damped transcripts (58.96% annotation, enrichment p-value=2.7e-10, and fold enrichment 1.566) and the chord diagram identified five highly connected genes: NCU00413, NCU00475, NCU00489, NCU00971, and NCU00706 (Figure 6A). The ribosome biogenesis GO term is also exclusively damped (Figure 6B), (57.53% annotation, enrichment p-value=6.7e-5, and fold enrichment 1.5283) but, unlike translation, the chord diagram does not reveal any genes with a large amount of connections. However, the shape of these protein-protein interactions are more interesting, dividing the genes largely into two sub-interacting groups based on phase (Figure 6B).

Finally, unlike the transcriptome, the *Neurospora* proteome was more evenly divided in categorization amongst AC groups. Interestingly, there was little enrichment in the damped or harmonic categories, with instead high amounts of enrichment in the forced category for metabolic categories. Our analysis found that the phosphorus metabolic process GO term was comprised exclusively of forced proteins (12.65% annotation, enrichment p-value=2.15e-2, and fold enrichment 1.5256) with two proteins that were highly connected, NCU09842 (MAK-1) and NCU02202 (Figure 6C).

4 DISCUSSION

We have above detailed a novel data analysis program, ENCORE. ENCORE bridges several different databases, including GO, UniProt, STRING, and QuickGo, to provide data visualizations that enable a deeper understanding to be drawn from omics-scale circadian data. ENCORE's Ontology Map subsets a large GO hierarchy to focus on key terms of interest while the Ontology Explorer provides numerical information of a specific term. The Group Comparison Tool

Figure 6: ENCORE demonstrates distinct differences between the rhythmic *Neurospora* proteome and transcriptome. Chord diagrams and mRNA levels of the highlighted genes corresponding to A. translation and B. ribosome biogenesis for the *Neurospora* transcriptome as well as a table of genes and their phases between the two gene groupings in B. C. Chord diagram and protein levels of the highlighted genes corresponding to phosphorus metabolic process for the *Neurospora* proteome. Data from [33].

displays STRING interactions for a selected GO term to highlight genes that may play an important role in that pathway's regulation. Further, the auxiliary information tab provides instant access to information about a gene of interest, making a comprehensive understanding of omics-data within reach. Finally, we have extended ENCORE to the Rensselaer Campfire, which reduces cognitive load and encourages collaboration, while engendering a new paradigm, Multi-Window Shiny, which extends the traditional Shiny framework into multiple connected windows [2, 19].

With ENCORE we demonstrated not only our ability to replicate previously published data but extend the findings of that work. For example, in Solenas et al. (2017) [49], ENCORE's Group Comparison Tool revealed several genes that were highly connected in the previously identified enriched GOs, including *Myc*, *Akt1*, *Mre11a*, *Rpa1*, and *Brca1* (Figure 4A and B). These genes are known to affect growth, cell death, and cancer, suggesting novel impacts of circadian regulation [15, 20, 37, 57, 60]. Meanwhile, ENCORE confirmed and emphasized the maintenance of rhythms in both adult and aged mice by highlighting the strong interconnection of the circadian clock network (Figure 4C).

These findings were bolstered when looking at Li et al. (2019) and Hurley et al. (2018); additionally in these data sets, ENCORE's ability to investigate the importance of AC categories was emphasized. ENCORE's inclusion of AC categories highlighted the vast differences between male/female flies and *Neurospora* transcriptome/proteome, respectively. For example, not only is it evident that the reproduction-associated genes *fru1* and *Gr2a* are key *Drosophila* regulators under the influence of the circadian clock, but the importance of a time-of-day regulation in courtship behavior can be extracted from the strong harmonic regulation that is seen in the genes in this pathway (Figure 5B) [39, 55, 56]. This is paralleled in the regulation of *Neurospora* translation and metabolic regulation, as genes in these GOs are all in particular AC categories.

ENCORE's visual capabilities also enabled the discovery of a phase of the day specific set of interacting groups in Hurley et al. (Figure 6B) ENCORE allowed for the recognition that in ribosomal biogenesis in *Neurospora*, daytime-present genes corresponded to ribonuclear proteins, while those in a group of night genes were exclusively ribosomal proteins. This emphasized that the circadian clock regulates the creation of key factors involved in the transport of transcriptional elements predominately during the day, while the creation of ribosomes occurs during the circadian evening.

Moreover, the ability of ENCORE to access UniProt allowed for the rapid identification of genes involved in many processes related to *Drosophila* and *Neurospora* development that were also determined to be clock-regulated (Figures 5A, 6) [6, 26, 35, 36]. Without ENCORE's ability to provide rapid, accessible visualizations based on omics-data, these connections may never have been achieved and therefore the development of ENCORE is likely to lead to the recognition of, and investigations into, many novel circadianly-regulated interactors.

In future work, we propose to add more ease of use to ENCORE by allowing gene-to-ontology search and added information to chords, such as simple expression correlation. We also seek to biologically investigate the hypotheses generated using the data from the novel ENCORE tool. As we demonstrated through our analysis of Hurley et al. (2018), ENCORE provides an impactful tool for the

comparison of different omics datasets. In our future work, we seek to extend this connection and build direct comparison into the tool itself. ENCORE's framework also provides an amenable space for additional analytics techniques, such as weighted correlation network analysis [61], which may provide further information depth to ENCORE's protein-protein interaction network. This amenability is emphasized by the fact that ENCORE is open-source and freely available, allowing for extensions on a variety of techniques and tools, such as the SWOT Clock for identifying windows of susceptibility [5]. These possible extensions also apply to different types of expression data, as well as different databases, such as KEGG and functional categories [38, 44]. ENCORE can also be applied to other rhythmic biological data, such as from the cell cycle field.

5 CONCLUSION

We present ENCORE, a novel tool for circadian biology which combines disparate databases for enhanced depth of understanding of circadian regulation. ENCORE's visualization techniques and inclusion of ECHO's AC categories engendered innovative insights into circadian gene set analysis, finding direct interactions between these categories and the prevalence of circadian regulation throughout biological functions. This is evidenced by our examination of previously published circadian studies spanning a diverse set of model organisms, emphasizing ENCORE's wide range. ENCORE satisfies an unmet need in the circadian community for ontological comparison and exploration of rhythmic genes.

ACKNOWLEDGMENTS

We would like to thank Emily Collins and Catherine Mann for their expertise in processing microarray data, Meaghan Jankowski for her expertise in *Neurospora*, and John Erickson for his campfire expertise. This work was supported by the National Institutes of Health (NIBIB U01 EB022546 to J.M.H. and H.D.I.S. and NIGMS R35 GM128687 to J.M.H.); the Department of Energy (PNNL 47818 to J.M.H.); Rensselaer Polytechnic Institute (to J.M.H. and H.D.I.S.); and the National Science Foundation (#1331023 to K.P.B.).

REFERENCES

- [1] Adrian Alexa and Jorg Rahnenfuhrer. 2016. *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.32.0.
- [2] Eric Louis Ameres. 2018. *Reducing the Cognitive Load of Visual Analytics of Networks Using Concentrically Arranged Multi-Surface Projections Focusing Immersive Real-Time Exploration*. Ph.D. Dissertation. Rensselaer Polytechnic Institute.
- [3] Michael Ashburner, et al. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* 25, 1 (may 2000), 25–29. <https://doi.org/10.1038/75556>
- [4] Sara J Aton, et al. 2005. Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. *Nature Neuroscience* 8, 4 (March 2005), 476–483. <https://doi.org/10.1038/nn1419>
- [5] Kristin P. Bennett, et al. 2019. Identifying Windows of Susceptibility by Temporal Gene Analysis. *Scientific Reports* 9, 1 (Dec. 2019), 2740. <https://doi.org/10.1038/s41598-019-39318-8>
- [6] Lindsay D. Bennett, et al. 2012. Circadian Activation of the Mitogen-Activated Protein Kinase MAK-1 Facilitates Rhythms in Clock-Controlled Genes in *Neurospora crassa*. *Eukaryotic Cell* 12, 1 (Nov. 2012), 59–69. <https://doi.org/10.1128/ec.00207-12>
- [7] D. Binns, et al. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25, 22 (sep 2009), 3045–3046. <https://doi.org/10.1093/bioinformatics/btp536>
- [8] Marc Carlson. 2018. *org.Ag.eg.db: Genome wide annotation for Anopheles*. R package version 3.7.0.
- [9] Marc Carlson. 2018. *org.Dm.eg.db: Genome wide annotation for Fly*. R package version 3.7.0.

- [10] Marc Carlson. 2018. *org.Eck12.eg.db: Genome wide annotation for E coli strain K12*. R package version 3.7.0.
- [11] Marc Carlson. 2018. *org.Hs.eg.db: Genome wide annotation for Human*. R package version 3.7.0.
- [12] Marc Carlson. 2018. *org.Mm.eg.db: Genome wide annotation for Mouse*. R package version 3.7.0.
- [13] Marc Carlson. 2018. *org.Sc.sgd.db: Genome wide annotation for Yeast*. R package version 3.7.0.
- [14] Winston Chang, et al. 2018. *shiny: Web Application Framework for R*. <https://CRAN.R-project.org/package=shiny> R package version 1.1.0.
- [15] W. S. Chen. 2001. Growth retardation and increased apoptosis in mice with homozygous disruption of the akt1 gene. *Genes & Development* 15, 17 (Sept. 2001), 2203–2208. <https://doi.org/10.1101/gad.913901>
- [16] Alexander M Crowell, et al. 2019. Learning and Imputation for Mass-spec Bias Reduction (LIMBR). *Bioinformatics* 35, 9 (May 2019), 1518–1526. <https://doi.org/10.1093/bioinformatics/bty828>
- [17] Hannah De los Santos, et al. 2017. Circadian Rhythms in Neurospora Exhibit Biologically Relevant Driven and Damped Harmonic Oscillations. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB '17)*. ACM, New York, NY, USA, 455–463. <https://doi.org/10.1145/3107411.3107420>
- [18] Hannah De los Santos, et al. 2019. ECHO: an Application for Detection and Analysis of Oscillators Identifies Metabolic Regulation on Genome-Wide Circadian Output. *bioRxiv* (2019). <https://doi.org/10.1101/690941> arXiv:<https://www.biorxiv.org/content/early/2019/07/03/690941.full.pdf>
- [19] Hannah De los Santos, et al. 2019. *mshiny: 'Shiny' for Multiple Windows*. <https://CRAN.R-project.org/package=mshiny> R package version 0.1.0.
- [20] R M Drost and J Jonkers. 2009. Preclinical mouse models for BRCA1-associated breast cancer. *British Journal of Cancer* 101, 10 (Sept. 2009), 1651–1657. <https://doi.org/10.1038/sj.bjc.6605350>
- [21] Jay C. Dunlap. 1999. Molecular bases for circadian clocks. [https://doi.org/10.1016/S0092-8674\(00\)80566-8](https://doi.org/10.1016/S0092-8674(00)80566-8) arXiv:362
- [22] Jay C. Dunlap and Jennifer J. Loros. 2017. Making Time: Conservation of Biological Clocks from Fungi to Animals. In *The Fungal Kingdom*. American Society of Microbiology, 515–534. <https://doi.org/10.1128/microbiolspec.funk-0039-2016>
- [23] Isaac Edery. 2000. Circadian rhythms in a nutshell. *Physiological Genomics* 3, 2 (Aug. 2000), 59–74. <https://doi.org/10.1152/physiolgenomics.2000.3.2.59>
- [24] Jennifer A. Evans and Alec J. Davidson. 2013. Health consequences of circadian disruption in humans and animal models. <https://doi.org/10.1016/B978-0-12-396971-2.00010-5> arXiv:arXiv:1011.1669v3
- [25] Andrea Franceschini, et al. 2012. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research* 41, D1 (Nov. 2012), D808–D815. <https://doi.org/10.1093/nar/gks1094>
- [26] James E. Galagan, et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422, 6934 (April 2003), 859–868. <https://doi.org/10.1038/nature01554>
- [27] Stacey L. Harmer, et al. 2001. Molecular Bases of Circadian Rhythms. *Annual Review of Cell and Developmental Biology* 17, 1 (Nov. 2001), 215–253. <https://doi.org/10.1146/annurev.cellbio.17.1.215>
- [28] Da Wei Huang, et al. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4, 1 (Jan. 2009), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- [29] Michael E Hughes, et al. 2017. Guidelines for Genome-Scale Analysis of Biological Rhythms. *Journal of Biological Rhythms* 32, 5 (oct 2017), 380–393. <https://doi.org/10.1177/0748730417728663>
- [30] Michael E. Hughes, et al. 2009. Harmonics of Circadian Gene Transcription in Mammals. *PLoS Genetics* 5, 4 (April 2009), e1000442. <https://doi.org/10.1371/journal.pgen.1000442>
- [31] Michael E. Hughes, et al. 2010. JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets. *Journal of Biological Rhythms* 25, 5 (Sept. 2010), 372–380. <https://doi.org/10.1177/0748730410379711>
- [32] Jennifer M. Hurley, et al. 2014. Analysis of clock-regulated genes in *Neurospora* reveals widespread posttranscriptional control of metabolic potential. *Proceedings of the National Academy of Sciences* 111, 48 (Dec. 2014), 16995–17002. <https://doi.org/10.1073/pnas.1418963111>
- [33] Jennifer M. Hurley, et al. 2018. Circadian Proteomic Analysis Uncovers Mechanisms of Post-Transcriptional Regulation in Metabolic Pathways. *Cell Systems* 7, 6 (Dec. 2018), 613–626.e5. <https://doi.org/10.1016/j.cels.2018.10.014>
- [34] Alan L. Hutchison, et al. 2015. Improved Statistical Methods Enable Greater Sensitivity in Rhythm Detection for Genome-Wide Data. *PLOS Computational Biology* 11, 3 (March 2015), e1004094. <https://doi.org/10.1371/journal.pcbi.1004094>
- [35] A. B. Jenkins. 2003. Drosophila E-cadherin is essential for proper germ cell-soma interaction during gonad morphogenesis. *Development* 130, 18 (Sept. 2003), 4417–4426. <https://doi.org/10.1242/dev.00639>
- [36] Jin Jiang and Gary Struhl. 1995. Protein kinase A and hedgehog signaling in drosophila limb development. *Cell* 80, 4 (Feb. 1995), 563–572. [https://doi.org/10.1016/0092-8674\(95\)90510-3](https://doi.org/10.1016/0092-8674(95)90510-3)
- [37] Yeun-Jin Ju, et al. 2006. Decreased expression of DNA repair proteins Ku70 and Mre11 is associated with aging and may contribute to the cellular senescence. *Experimental & Molecular Medicine* 38, 6 (Dec. 2006), 686–693. <https://doi.org/10.1038/emmm.2006.81>
- [38] M. Kanehisa. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28, 1 (Jan. 2000), 27–30. <https://doi.org/10.1093/nar/28.1.27>
- [39] Haein Kim, et al. 2017. Involvement of a Gr2a-Expressing Drosophila Pharyngeal Gustatory Receptor Neuron in Regulation of Aversion to High-Salt Foods. *Molecules and Cells* 40, 5 (2017), 331–338. <https://doi.org/10.14348/molcells.2017.0028>
- [40] Jiajia Li, et al. 2019. Achilles-Mediated and Sex-Specific Regulation of Circadian mRNA Rhythms in Drosophila. *Journal of Biological Rhythms* 34, 2 (Feb. 2019), 131–143. <https://doi.org/10.1177/0748730419830845>
- [41] Javier Luraschi and JJ Allaire. 2018. *r2d3: Interface to 'D3' Visualizations*. <https://CRAN.R-project.org/package=r2d3> R package version 0.2.3.
- [42] Hervé Pagès, et al. 2018. *AnnotationDbi: Annotation Database Interface*. R package version 1.44.0.
- [43] Carrie L. Partch, et al. 2014. Molecular architecture of the mammalian circadian clock. <https://doi.org/10.1016/j.tcb.2013.07.002> arXiv:NIHMS150003
- [44] Steffen Priebe, et al. 2011. FungiFun: A web-based application for functional categorization of fungal genes and proteins. *Fungal Genetics and Biology* 48, 4 (apr 2011), 353–358. <https://doi.org/10.1016/j.fgb.2010.11.001>
- [45] R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- [46] Alexander Roth, et al. 2016. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* 45, D1 (10 2016), D362–D368. <https://doi.org/10.1093/nar/gkw937> arXiv:<http://oup.prod.sis.lan/nar/article-pdf/45/D1/D362/8847225/gkw937.pdf>
- [47] S. S. C. Rund, et al. 2011. Genome-wide profiling of diel and circadian gene expression in the malaria vector *Anopheles gambiae*. *Proceedings of the National Academy of Sciences* 108, 32 (2011), E421–E430. <https://doi.org/10.1073/pnas.1100584108> arXiv:arXiv:1408.1149
- [48] M. L. Sargent, et al. 1966. Circadian Nature of a Rhythm Expressed by an Invertaseless Strain of *Neurospora crassa*. *PLANT PHYSIOLOGY* 41, 8 (Oct. 1966), 1343–1349. <https://doi.org/10.1104/pp.41.8.1343>
- [49] Guiomar Solanas, et al. 2017. Aged Stem Cells Reprogram Their Daily Rhythmic Functions to Adapt to Stress. *Cell* 170, 4 (aug 2017), 678–692.e20. <https://doi.org/10.1016/j.cell.2017.07.035>
- [50] A. Subramanian, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102, 43 (Sept. 2005), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- [51] Fran Supek, et al. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6, 7 (July 2011), e21800. <https://doi.org/10.1371/journal.pone.0021800>
- [52] P. D. Thomas. 2003. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research* 13, 9 (sep 2003), 2129–2141. <https://doi.org/10.1101/gr.772403>
- [53] M. Tory and T. Moller. 2004. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics* 10, 1 (Jan. 2004), 72–84. <https://doi.org/10.1109/tvcg.2004.1260759>
- [54] The Uniprot Consortium. 2018. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* 47, D1 (nov 2018), D506–D515. <https://doi.org/10.1093/nar/gky1049>
- [55] Adriana Vilella, et al. 1997. Extended Reproductive Roles of the fruitless Gene in Drosophila melanogaster Revealed by Behavioral Analysis of New fru Mutants. *Genetics* 147, 3 (1997), 1107–1130. arXiv:<https://www.genetics.org/content/147/3/1107.full.pdf> <https://www.genetics.org/content/147/3/1107>
- [56] Eleftheria Vrontou, et al. 2006. fruitless regulates aggression and dominance in Drosophila. *Nature Neuroscience* 9, 12 (Nov. 2006), 1469–1471. <https://doi.org/10.1038/nn1809>
- [57] Yuxun Wang, et al. 2005. Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. *Nature Genetics* 37, 7 (June 2005), 750–755. <https://doi.org/10.1038/ng1587>
- [58] S. Yamazaki. 2000. Resetting Central and Peripheral Circadian Oscillators in Transgenic Rats. *Science* 288, 5466 (April 2000), 682–685. <https://doi.org/10.1126/science.288.5466.682>
- [59] Jia Ye, et al. 2018. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Research* 46, W1 (July 2018), W71–W75. <https://doi.org/10.1093/nar/gky400>
- [60] Jian Yuan, et al. 2009. A c-Myc-SIRT1 feedback loop regulates cell growth and transformation. *The Journal of Cell Biology* 185, 2 (April 2009), 203–211. <https://doi.org/10.1083/jcb.200809167>
- [61] Bin Zhang and Steve Horvath. 2005. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4, 1 (2005).