

RESEARCH ARTICLE

# Assembly of a Comprehensive Regulatory Network for the Mammalian Circadian Clock: A Bioinformatics Approach

Robert Lehmann<sup>1</sup>, Liam Childs<sup>2</sup> , Philippe Thomas<sup>2</sup> , Monica Abreu<sup>1,3</sup>, Luise Fuhr<sup>1,3</sup>, Hanspeter Herzel<sup>1</sup>, Ulf Leser<sup>2</sup>, Angela Relógio<sup>1,3\*</sup>

**1** Institute for Theoretical Biology (ITB), Charité-Universitätsmedizin Berlin and Humboldt-Universität zu Berlin, Invalidenstraße 43, 10115, Berlin, Germany, **2** Knowledge Management in Bioinformatics, Institute for Computer Science, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany, **3** Molekulares Krebsforschungszentrum (MKFZ), Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353, Berlin, Germany

 These authors contributed equally to this work.

\* [angela.relogio@charite.de](mailto:angela.relogio@charite.de)



 OPEN ACCESS

**Citation:** Lehmann R, Childs L, Thomas P, Abreu M, Fuhr L, Herzel H, et al. (2015) Assembly of a Comprehensive Regulatory Network for the Mammalian Circadian Clock: A Bioinformatics Approach. PLoS ONE 10(5): e0126283. doi:10.1371/journal.pone.0126283

**Academic Editor:** Nicolas Cermakian, McGill University, CANADA

**Received:** January 19, 2015

**Accepted:** March 31, 2015

**Published:** May 6, 2015

**Copyright:** © 2015 Lehmann et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was funded by the BMBF (eBio-CIRSPICE grant to AR LF MA, eBio-OncoPath grant to LS, UL) and graduate school SOAMED (PT), the Berlin School of Integrative Oncology (BSIO) of Charité Universitätsmedizin Berlin (LF) and the Deutsche Forschungsgemeinschaft (SFB 618/ A4). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

By regulating the timing of cellular processes, the circadian clock provides a way to adapt physiology and behaviour to the geophysical time. In mammals, a light-entrainable master clock located in the suprachiasmatic nucleus (SCN) controls peripheral clocks that are present in virtually every body cell. Defective circadian timing is associated with several pathologies such as cancer and metabolic and sleep disorders. To better understand the circadian regulation of cellular processes, we developed a bioinformatics pipeline encompassing the analysis of high-throughput data sets and the exploitation of published knowledge by text-mining. We identified 118 novel potential clock-regulated genes and integrated them into an existing high-quality circadian network, generating the to-date most comprehensive network of circadian regulated genes (NCRG). To validate particular elements in our network, we assessed publicly available ChIP-seq data for BMAL1, REV-ERBa/β and RORα/γ proteins and found strong evidence for circadian regulation of *Elavl1*, *Nme1*, *Dhx6*, *Med1* and *Rbbp7* all of which are involved in the regulation of tumourigenesis. Furthermore, we identified *Ncl* and *Ddx6*, as targets of RORγ and REV-ERBa, β, respectively. Most interestingly, these genes were also reported to be involved in miRNA regulation; in particular, NCL regulates several miRNAs, all involved in cancer aggressiveness. Thus, NCL represents a novel potential link via which the circadian clock, and specifically RORγ, regulates the expression of miRNAs, with particular consequences in breast cancer progression. Our findings bring us one step forward towards a mechanistic understanding of mammalian circadian regulation, and provide further evidence of the influence of circadian deregulation in cancer.

**Competing Interests:** The authors have declared that no competing interests exist.

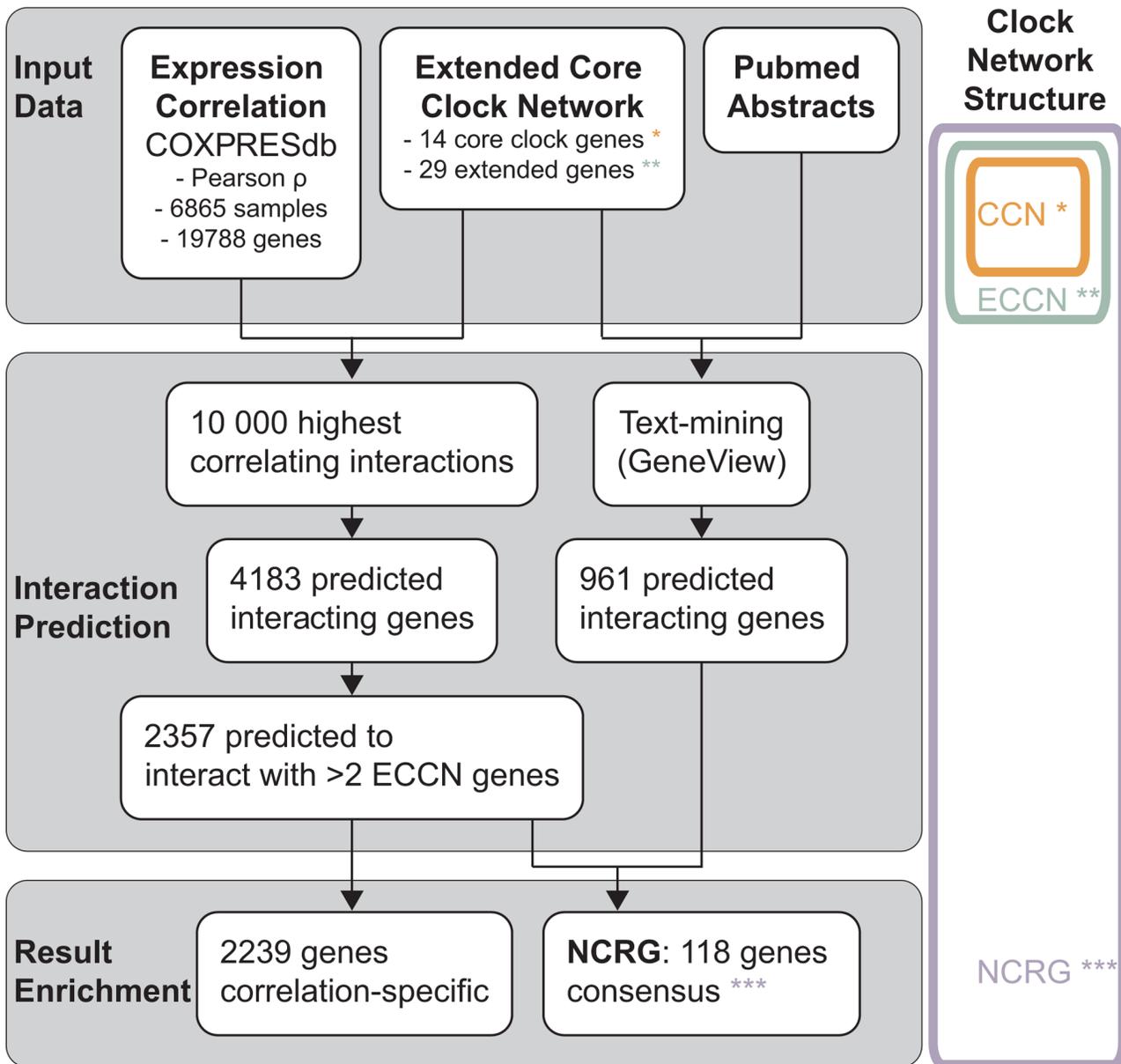
## Introduction

Almost all organisms evolved an endogenous circadian clock which regulates the timing of central biological processes and provides a way to adapt physiology and behaviour to daily dark/light rhythms [1–3]. In mammals, malfunctions of the circadian system are associated to known pathologies ranging from sleep or metabolic disorders, to cancer [4–6]. Hence, a detailed overview of the underlying genetic network that shapes the mammalian circadian system is of major interest to the circadian and medical field.

The mammalian circadian system is hierarchically organized. A main pacemaker formed by two clusters of ~100,000 neurons (in humans) is located in the suprachiasmatic nucleus (SCN), but peripheral oscillators exist in virtually every of our  $3.5 \times 10^{13}$  body cells [7, 8]. Extensive research has identified a reduced set of 14 genes to form the so called core-clock network (CCN), within a cell. These genes encode for members of several gene families: PER (period), CRY (cryptochrome), BMAL (brain and muscle ARNT-like protein), CLOCK (circadian locomotor output cycles kaput), NPAS2 (neuronal PAS domain-containing protein 2, in neuronal tissue), ROR (retinoic acid receptor-related orphan receptor) and REV-ERB (nuclear receptor, reverse strand of ERBA). The CCN is arranged in two main interconnected feed-back loops: a) the RORs/*Bmal*/REV-ERBs (RBR) loop and b) the PERs/CRYs (PC) loop [9]. Both loops are able to produce rhythms in gene expression, independently, but need to be interconnected to robustly generate oscillations with a period of circa 24 hours [10, 11]. In the centre of the core-clock network lays the heterodimer complex CLOCK/BMAL1. This complex regulates the transcription of elements of both the RBR and PC loop by binding to E-Box sequences in the promoter region of the target genes. In the RBR-loop, *Rev-Erba*, $\beta$  and *Rora*, $\beta$ , $\gamma$  are transcribed. After translation, the resulting proteins compete for RORE elements within the *Bmal1* promoter region and hold antagonistic effects, thereby fine-tuning *Bmal1* expression. In the PC loop, following transcription and translation, PER1,2,3 and CRY1,2 form complexes and inhibit CLOCK/BMAL mediated-transcription, thus regulating the expression of all target genes mentioned above.

The CCN has been studied, on a fine scale, at the transcriptional, translational and post-translational level both experimentally and with mathematical models [9, 12–18]. Furthermore, various efforts have been made to decipher the mechanisms through which the mammalian CCN regulates its target genes, the clock-controlled genes (CCG), as well as to identify new CCGs [19–21]. Yet, a more detailed knowledge on the full range of genes and subsequent biological processes that are regulated by the core of the circadian clock is still missing. Therefore, a comprehensive analysis of the relevance of such connections, as well as on the putative effects of deregulations on circadian output and resulting pathological phenotypes, is needed.

In this manuscript, we present a comprehensive mammalian circadian network constructed by an integrated bioinformatics pipeline which uses different data sources and different data types. This novel circadian network topology highlights particularly genes which link the circadian clock to several biological processes, often in multiple alternative ways. We carried out a systematic expansion of a previously published core-clock network (ECCN) using gene co-expression analysis, text-mining on the full PubMed, signatures of circadian expression patterns, and CHIP-Seq data. We used the first two of these methods to identify a set of 118 novel high-confidence ECCN target genes, whereas the latter two data types were used for validation of this set, which resulted in a novel network of circadian regulated genes (NCRG) (Fig 1). In particular, CHIP-seq data for BMAL1, ROR $\alpha$ , $\gamma$  and REV-ERB $\alpha$ , $\beta$  [12, 15–17, 22] confirmed links between the ECCN and several cancer-related genes. Notably, two of these genes were shown to be involved in miRNA regulation.



**Fig 1. Work flow used to establish a network of circadian regulated genes (NCRG).** Two independent data types were used to predict genes which interact with the human core clock network (CCN, orange) and the extended core clock network (ECCN, green). Co-expression data was used to find sets of genes with strongest (anti-) correlating expression with the 43 ECCN genes across a large number of independent experiments. A total of 2357 genes were found to interact with more than 2 ECCN genes. The GeneView text-mining pipeline was used to analyse published knowledge (approximately 22 million citations) about interacting genes. A total of 961 text-mining-predicted genes were found. The intersection of both methodologies resulted in 118 new genes, which together with the ECCN form a new network of circadian regulated genes (NCRG, purple).

doi:10.1371/journal.pone.0126283.g001

Altogether, our findings suggest, new potential clock genes and describe their role and topology within the circadian network. Our work delivers novel evidence to the influence of circadian deregulation in cancer and adds a novel way via which a clock-dependent cancer output may emerge, i.e., miRNA circadian regulation.

## Results

### A text-mining based approach for network discovery

We aimed to update and extend our recently reported circadian network (ECCN-extended core-clock network) [21]. For that we combined the results of a text-mining system with high-throughput gene co-expression data to obtain new elements and interactions. This procedure resulted in a network of circadian regulated genes (NCRG) following the workflow schematized in Fig 1.

The original ECCN [21] contains a core of 14 well known circadian genes, *Per1,2,3*, *Cry1,2*, *Bmal1,2*, *Rora,β,γ* and *Rev-Erba,β* as well as *Clock*, its paralog *Npas2*, and their direct neighbouring targets. We started our study by generating an update version of the ECCN using the text-mining software—GeneView (see [Materials and Methods](#)) [23] to extract all pairwise interactions among our genes of interest and their directly interacting neighbours. The new ECCN contains 43 elements as the previous network [21], the depicted interactions were updated to the current PubMed available data resulting in more than 200 regulatory relationships (Fig 2). Additional information containing all interactions and corresponding references, as well as a more detailed characterization of the ECCN is provided in [S1 Table](#) and [S1 Text](#), respectively.

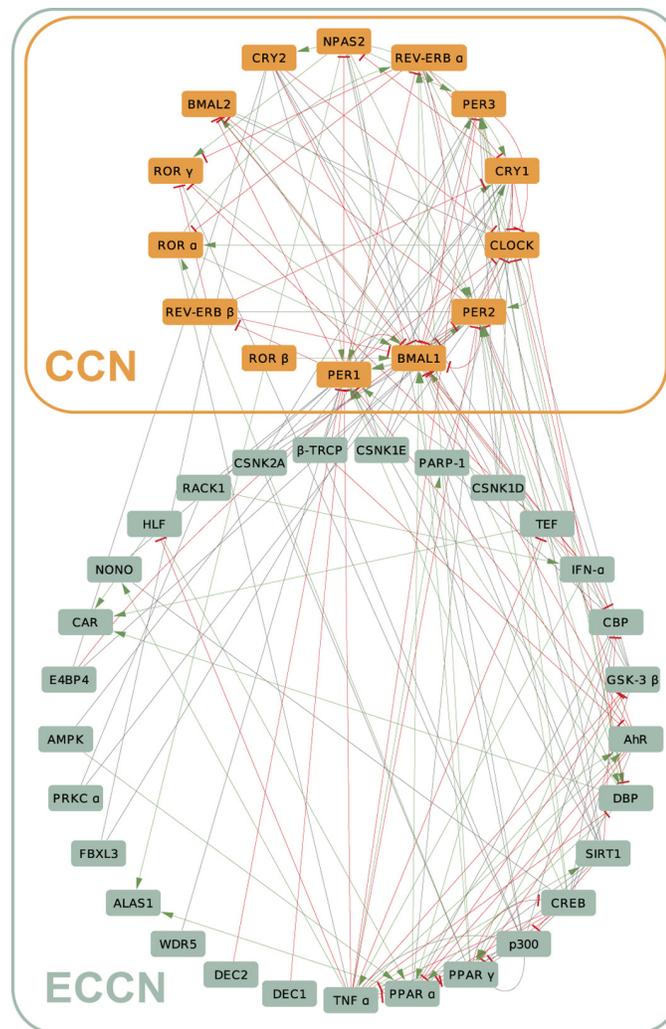
### Co-expression data analysis confirms the ECCN network topology

In this work we expanded the updated ECCN with a new layer of potentially ECCN-regulated elements (genes and proteins) using co-expression data as a first source of evidence. We consider as such candidates all genes which show a strong co-expression to ECCN members and which can be confirmed using text-mining, as indicated in Fig 1.

There are several public available databases providing co-expression metrics for human genes, we evaluated four different such databases [24–26] regarding their ability to reproduce the ECCN, to find the best suited for our analysis. Results of our comparisons are presented in [S2 Text](#). We eventually choose COXPRESdb [25] (see [Fig 3](#)), as this database showed the highest degree of correlation within all genes of the extended core-clock network.

We expected a significant difference between the correlation measure distributions of interacting gene pairs and a chosen background, if unknown interacting pairs were to be predicted based on correlation values. All possible pairs between a random set of 43 genes and all genes (19,788) were used as background. As foreground pairs we used the 42 known available interactions amongst the CCN genes, as well as the 119 curated interactions of the ECCN gene set. Both the CCN (orange) and ECCN (green) gene pairs tended to have higher correlations compared to the random background and thus lower mutual rank (MR) values (Fig 3A and 3B and [S1 Fig](#)). The probability density functions of correlation and mutual rank are shown in [S2 Fig](#) for both datasets. All correlation values were Fisher transformed to ensure normal distribution prior to hypothesis testing to characterize differences between the CCN, ECCN, and the background. Subsequent one-sided t-tests with the alternative hypothesis to observe smaller correlations in the background confirmed the results of the visual inspection: CCN gene pairs are significantly higher correlated than the background pairs ( $p < 0.0195$ ) similar to the ECCN gene pairs ( $p < 6e^{-7}$ ). Similarly significant differences were observed in the Hsa dataset for the CCN ( $p < 1.4e^{-3}$ ) and the ECCN ( $p < 1.2e^{-7}$ ). Furthermore, no significant difference was found between correlations in the CCN and the ECCN for the Hsa ( $p < 0.41$ ) and the Hsa2 dataset ( $p < 0.18$ ).

Studying the co-expression of ECCN genes in detail, we observed anti-correlated circadian expression profiles between *Per* and *Bmal*, which is consistent with the predicted 9h delay

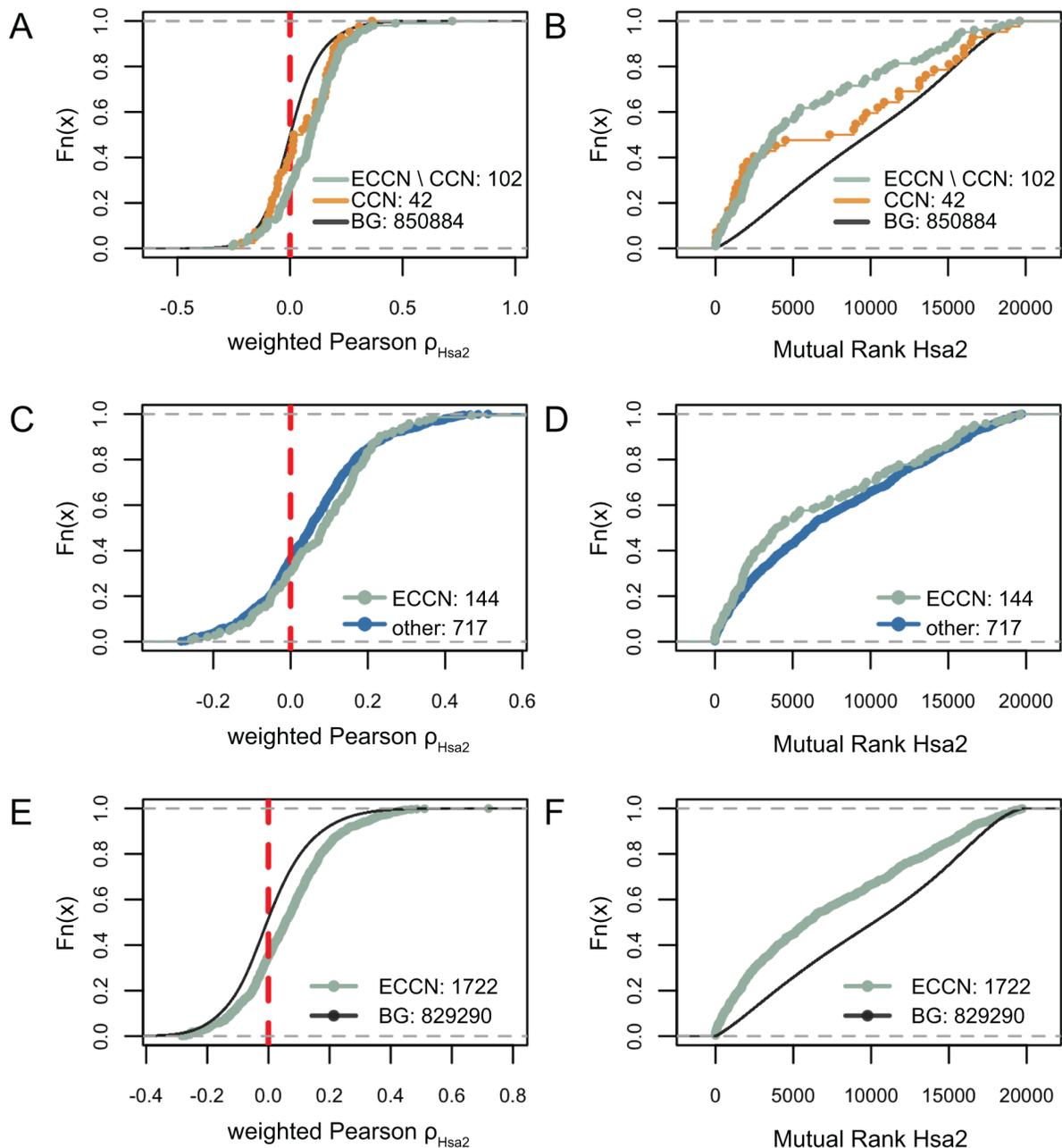


**Fig 2. The human core clock network (CCN) and the extended core clock network (ECCN).** The CCN (orange) contains the known core-clock elements (Per1,2,3, Cry1,2, Rev-Erb $\alpha,\beta$ , Rora, $\beta,\gamma$ , Bmal1,2, Clock and Npas). The ECCN (green) was obtained after an extensive collection of CCN-interacting genes followed by a detailed curation for direct interactions [21] and a further update to the recent literature. Activation (green lines), inhibition (red lines) and other sort of interactions (grey lines) are represented. The resulting clock network contains 43 elements and more than 200 regulatory relationships.

doi:10.1371/journal.pone.0126283.g002

between the mRNAs peak expression for these genes [9]. The strongest anti-correlation was determined between  $\rho_{\text{HSA}_2}(\text{Bmal1}, \text{Per3}) = -0.21$  and the weakest between  $\rho_{\text{HSA}_2}(\text{Bmal2}, \text{Per2}) = -0.054$ . For the expected expression-correlation between *Ror* and *Rev-Erb* (circa 7h delay), we determined weaker anti-correlations, with  $\rho_{\text{HSA}_2}(\text{Rev-Erb}\beta, \text{Ror}\gamma) = -0.04$ ,  $\rho_{\text{HSA}_2}(\text{Rev-Erb}\alpha, \text{Ror}\gamma) = 0.06$ , and  $\rho_{\text{HSA}_2}(\text{Rev-Erb}\alpha, \text{Ror}\alpha) = 0.16$ . We could also validate the expected positive correlation between *Per* and *Cry*, specifically for *Per2/Cry2* with  $\rho_{\text{HSA}_2}(\text{Per2}, \text{Cry2}) = 0.31$ . For other *Per* and *Cry* family members, smaller correlation values were found with  $\rho_{\text{HSA}_2}(\text{Per2}, \text{Cry1}) = 0.12$ , and  $\rho_{\text{HSA}_2}(\text{Per1}, \text{Cry1}) = 0.0022$ .

Next, we tested whether the distribution of correlations between pairs of reported interacting ECCN genes could be distinguished from all other possible pairs of ECCN genes. The known interactions exhibited positive  $\rho_{\text{HSA}_2}$  values and thus lower MR (Fig 3C and 3D). However, only a weak tendency of known interactions towards higher correlations compared to



**Fig 3. Correlation distributions for clock network gene pairs versus random gene pairs.** The cumulative Pearson  $\rho$  distributions of pairs of ECCN genes reported to interact but excluding CCN (ECCN, green), reported pairs of CCN genes (CCN, orange), and 43 randomly chosen genes versus all genes as background (BG, black) are shown for the Hsa2 data collection (A, B). Distributions are shown centred around 0 with the centred bin marked by the dashed red line. The Pearson  $\rho$  distributions of reported pairs of ECCN genes (green) is compared to not reported pairs (blue) for the data set Hsa2 (C, D). Comparison of Pearson  $\rho$  and mutual rank (E, F) between all possible pairs of ECCN genes (green) and all possible pairs of non-ECCN genes as background (black). All data were taken from the Hsa2 data collection.

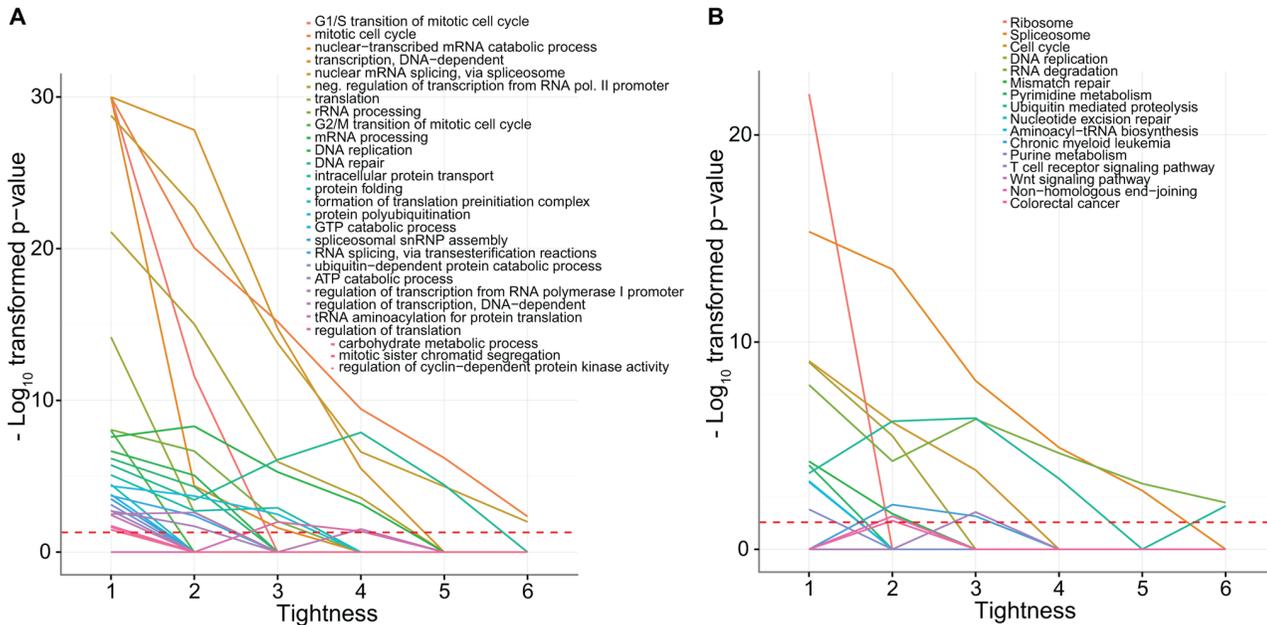
doi:10.1371/journal.pone.0126283.g003

non-reported pairs can be observed (probability density functions shown in S3 Fig). The corresponding comparisons of correlation distributions via t-tests confirmed the weak tendency in the Hsa2 dataset ( $p < 0.059$ ) whereas no signal was found in the Hsa dataset ( $p < 0.395$ ). As a consequence, we conclude that known interacting gene pairs cannot reliably be distinguished

from other pairs within the ECCN based merely on expression correlation. We then tested the assumption that expression patterns are generally higher correlated within the ECCN as compared to other non-ECCN gene pairs. The resulting distributions for Pearson  $\rho$  and mutual rank between all possible pairs of the ECCN genes included in the datasets, as compared to all combinations of the same genes with all other genes are shown in Fig 3E and 3F (probability density functions shown in S4 Fig). We confirmed the difference of the ECCN set as compared to the background set as before with t-test which yielded high significance in both datasets ( $p < 2.2e^{-16}$ ). In addition, the corresponding mutual rank measures were also found to be significantly lower (Wilcoxon Rank Sum test,  $p < 2.2e^{-16}$ ). Hence, we concluded that the examined expression correlation data provided information about the membership of a gene to the clock network, but not about the network's topology.

### Expression correlation-based target prediction

We selected the 10,000 highest correlating pairs of one of the 43 ECCN genes and any other gene. This conservatively chosen threshold selects 1.18% of all pairs, corresponding to an absolute correlation cut-off of 0.3636, or  $2.6\sigma$ . In this set, the number of unique new genes was 4,183. As we sought to investigate genes that were tightly associated with the ECCN (i.e. associated with multiple ECCN genes). We defined tightness as the number of connections between a gene and the ECCN and sought to find the largest number of tightly connected predicted targets by filtering them at several levels of minimum tightness. At increasing levels of minimum tightness, we performed an overrepresentation analysis of GO and KEGG terms for varying minimal values of these counts. The largest changes in overrepresented terms occurred when changing the threshold from one to two (Fig 4). There, terms related to cell cycle and the



**Fig 4. Variation in functional annotation enrichment with increasing tightness between the predicted targets and the ECCN.** The significance of 28 enriched GO terms (A) and 16 KEGG terms (B) for genes connected to the CCN steadily decreases for most terms as the minimum number connections is increased (this number of connections between a gene and a gene set is here defined as "tightness"). As the minimum tightness between the predicted targets and the ECCN increases, the enrichment and rank of functional annotation changes. We observe an overall decrease in enrichment but little in rank. The greatest changes in rank occur between a tightness of 2 and 3. At a tightness of two and above, the rank of the majority of significant GO terms such as "mitotic cell cycle" and "nuclear mRNA splicing, via spliceosome", and KEGG terms such as "Spliceosome", "Ubiquitin mediated proteolysis" and "RNA degradation" remain largely stable suggesting a natural threshold on tightness at this point.

doi:10.1371/journal.pone.0126283.g004

ribosome rapidly dropped in significance, while, terms related to splicing and transcription largely retained their position. When increasing the tightness further, much smaller changes occurred.

Accordingly, we defined tightly connected genes as those having two or more associations with the ECCN, which reduced the number of predicted ECCN targets from 4,183 to 2,357 with 8,180 interactions (S5 Fig).

The number of tightly connected genes associated to a given ECCN element varied greatly. While 11 ECCN members did not feature any interaction, the three elements CREB, AMPK, and CLOCK covered 48% of all predicted interactions (S6 Fig).

Gene ontology terms (GO) and KEGG pathway enriched in this set are listed in Table 1 and Table 2, respectively. To obtain an insight into which ECCN genes are of particular importance for the enriched function or pathway, we determined the cross table for each combination of ECCN gene versus enriched term, counting the number of predicted target genes featuring the corresponding term. This approach yielded a consistent pattern for GO functional annotations (50 terms with  $q < 0.01$ ) and KEGG pathways (35 pathways with  $q < 0.01$ ) (Fig 5). About half of the ECCN genes were associated with a multitude of genes covering a range of GO annotations, while the other half was associated with few genes covering only a small number of GO terms (Fig 5A). The largest number of target genes was annotated with the molecular function “protein binding” and the cellular component “nucleus”. The second-strongest molecular function signal was “DNA binding”. The most striking association was found between the genes *Csnk2a*, *Wdr5*, *Nono*, and *Parp-1* and the spliceosome ( $q < 1.5e^{-37}$ ) and RNA transport ( $q < 8e^{-38}$ ) pathways, where  $q$  represent the p-value adjusted by Benjamini-Hochberg multiple testing correction. These genes were also predicted to target ribosome biogenesis, cell cycle, and purine/pyrimidine synthesis related genes. Another strong association was found between cancer-related pathways such as “Pathways in cancer” ( $q < 7e^{-7}$ ), “Wnt signalling” ( $q < 2.8e^{-8}$ ), “MAPK signalling” ( $q < 4e^{-6}$ ), and the *Ampk* and *Creb* target genes (Fig 5B).

## An extended network of circadian regulation: beyond the core

We used text-mining to obtain a second set of genes potentially regulated by the ECCN, and then compared this set to the 2357 genes obtained from co-expression analysis (Fig 1). First, we obtained from GeneView the 50 most frequent interaction partners for each ECCN element, resulting in 961 new interacting genes, each supported by 55 sentences on average. These genes and their supporting sentences are given in S2 Table. The analysis of a large set of GeneView-output sentences revealed 20% of wrong sentences which corresponded to 10% false-positive interactions. Again, we subjected this gene set to enrichment analysis. A large number of significantly enriched annotations were observed in the analysis of GO terms (154 terms with  $q < 0.01$ ) and KEGG pathways (115 pathways with  $q < 0.01$ ) (S3 and S4 Tables). The top 4 GO terms ( $q < 7.6e^{-18}$ ) included positive and negative regulation of transcription from RNA polymerase II promoters (GO:0045944, GO:0000122), indicating a large fraction of transcription regulatory genes in this set. The term “anti-apoptosis” was listed on the 7<sup>th</sup> position ( $q < 7.5e^{-13}$ ) with 64 annotations found, where only 16 are expected by chance. The top-three enriched KEGG annotations were “Pathways in cancer” ( $q < 1.6e^{-92}$ ), “Cytokine-cytokine receptor interaction” ( $q < 3e^{-34}$ ), and “Toll-like receptor signalling pathway” ( $q < 2.6e^{-35}$ ), with a range of cancer-related pathways following.

Intersecting the ECCN-interacting gene sets predicted by expression correlation ( $n = 2357$ ) and text-mining ( $n = 961$ ), respectively, resulted in a set of 118 genes (Fig 6A). While 38 novel interactions with an ECCN gene were predicted by both methods, 364 interactions were co-expression-specific and 182 were text-mining-specific (S5 Table). Interestingly, enrichment

Table 1. Enrichment analysis of the co-expression-predicted ECCN interacting genes for GO term annotations.

GO Term	Annotations in Total	Annotations in Predicted Set	Expected	FDR
nuclear mRNA splicing, via spliceosome	200	88	27	1.9e-10
cell division	443	114	59.81	1.6e-09
mRNA transport	105	54	14.18	6.7e-08
DNA strand elongation involved in DNA replication	34	21	4.59	1.2e-06
ubiquitin-dependent protein catabolic process	347	104	46.85	1.7e-06
S phase of mitotic cell cycle	130	44	17.55	1.8e-06
regulation of glucose transport	69	24	9.32	1.9e-06
mitotic prometaphase	84	35	11.34	1.9e-06
M/G1 transition of mitotic cell cycle	76	32	10.26	8.4e-06
mRNA processing	378	152	51.03	1.9e-05
gene expression	4347	874	586.86	0.00027
cell cycle checkpoint	234	70	31.59	0.00031
DNA duplex unwinding	28	18	3.78	0.00033
nuclear-transcribed mRNA poly(A) tail shortening	26	16	3.51	0.00038
mRNA export from nucleus	59	28	7.97	0.00042
protein transport	1154	222	155.79	0.00047
DNA-dependent DNA replication initiation	28	16	3.78	0.00079
DNA repair	369	109	49.82	0.00120
termination of RNA polymerase II transcription	44	20	5.94	0.00285
regulation of transcription, DNA-templated	2735	506	369.23	0.00518
RNA splicing	308	121	41.58	0.00989
protein binding	6831	1116	889.85	2.6e-29
RNA binding	792	248	103.17	7.0e-24
DNA binding	2240	454	291.8	5.8e-14
ATP binding	1439	292	187.45	9.4e-13
nucleotide binding	2294	448	298.83	3.3e-07
ubiquitin thiolesterase activity	64	27	8.34	2.6e-05
chromatin binding	235	59	30.61	0.00026
ubiquitin-protein ligase activity	241	63	31.39	0.00104
translation initiation factor activity	50	21	6.51	0.00134
ubiquitin-specific protease activity	43	19	5.6	0.00186
ATP-dependent DNA helicase activity	32	17	4.17	0.00547
protein transporter activity	87	28	11.33	0.00865
nucleus	5640	1172	724.54	2.1e-30
nucleoplasm	1401	423	179.98	1.0e-27
nuclear speck	144	61	18.5	1.3e-15
nucleolus	589	154	75.67	2.4e-14
catalytic step 2 spliceosome	78	40	10.02	3.7e-13
nuclear pore	60	36	7.71	5.6e-10
cytosol	2217	382	284.8	3.6e-07
heterogeneous nuclear ribonucleoprotein complex	19	14	2.44	2.5e-06
centrosome	363	89	46.63	3.9e-05
Cajal body	44	20	5.65	0.00014
cytoplasmic stress granule	21	12	2.7	0.00239
spliceosomal complex	137	60	17.6	0.00322
nuclear pore outer ring	10	8	1.28	0.00330
chromatin	280	72	35.97	0.00466

(Continued)

Table 1. (Continued)

GO Term	Annotations in Total	Annotations in Predicted Set	Expected	FDR
DNA replication factor C complex	6	6	0.77	0.00568
chaperonin-containing T-complex	6	6	0.77	0.00568
nuclear membrane	169	46	21.71	0.00685

Gene ontology annotation enrichment was performed for the molecular function, cellular component, and biological process ontologies. Only terms with  $q < 0.01$  (false discovery rate after Benjamini-Hochberg) are shown.

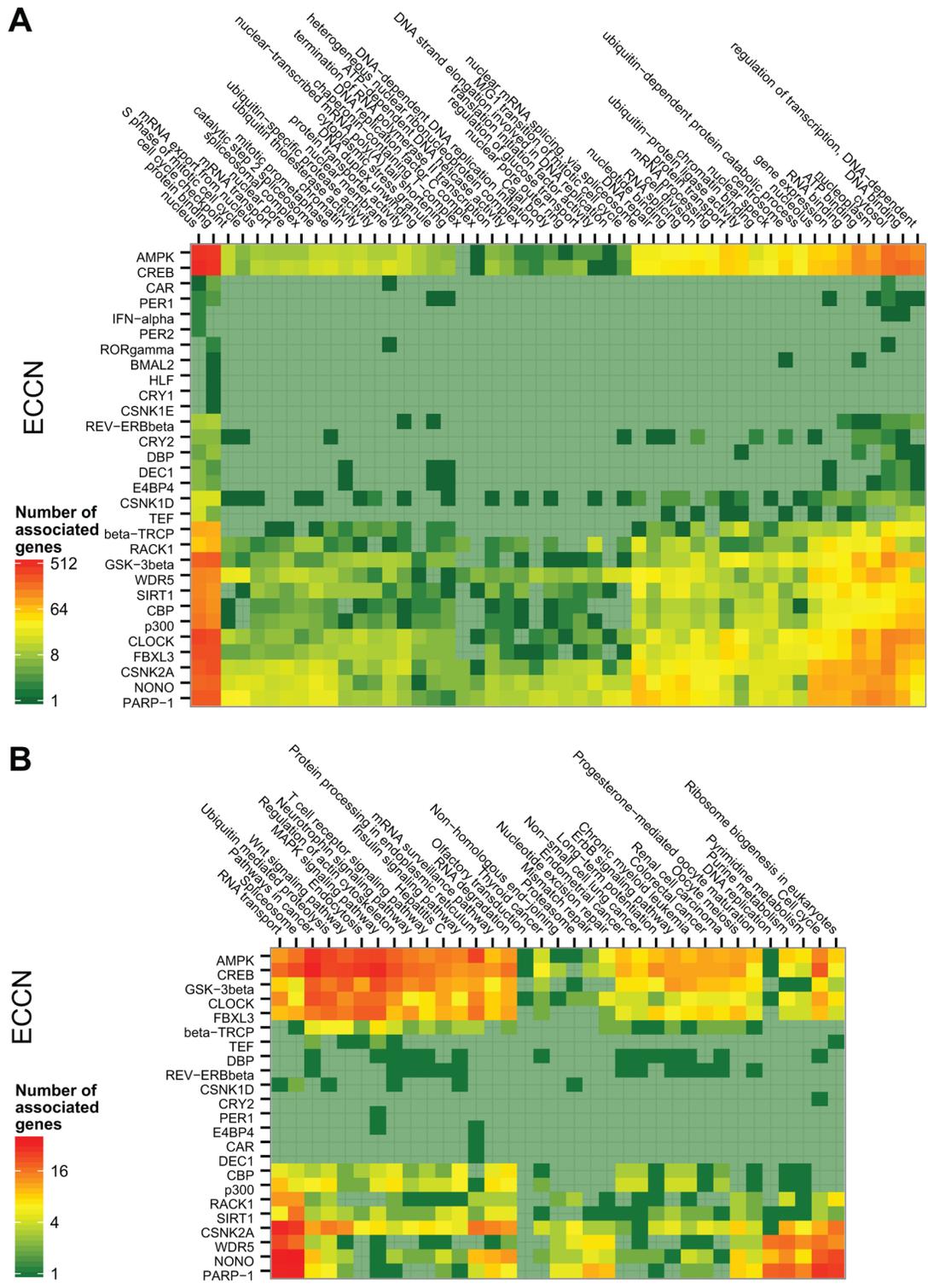
doi:10.1371/journal.pone.0126283.t001

Table 2. Enrichment of KEGG pathway annotations amongst the co-expression-predicted ECCN interacting genes.

KEGG ID	Pathway	p-value	FDR
hsa03013	RNA transport	3.51E-40	8.00E-38
hsa03040	Spliceosome	6.64E-40	1.50E-37
hsa04110	Cell cycle	7.95E-23	1.80E-20
hsa03008	Ribosome biogenesis in eukaryotes	6.98E-21	1.60E-18
hsa04120	Ubiquitin mediated proteolysis	4.01E-16	9.00E-14
hsa03018	RNA degradation	1.95E-14	4.40E-12
hsa03030	DNA replication	6.08E-14	1.40E-11
hsa03015	mRNA surveillance pathway	8.00E-12	1.80E-09
hsa04141	Protein processing in endoplasmic reticulum	2.52E-11	5.60E-09
hsa04310	Wnt signalling pathway	1.26E-10	2.80E-08
hsa05200	Pathways in cancer	3.19E-09	7.00E-07
hsa04740	Olfactory transduction	6.69E-09	1.50E-06
hsa03430	Mismatch repair	8.92E-09	1.90E-06
hsa00230	Purine metabolism	1.24E-08	2.70E-06
hsa00240	Pyrimidine metabolism	1.24E-08	2.70E-06
hsa04010	MAPK signalling pathway	1.84E-08	3.90E-06
hsa03420	Nucleotide excision repair	6.15E-08	1.30E-05
hsa05220	Chronic myeloid leukemia	2.60E-07	5.50E-05
hsa04722	Neurotrophin signalling pathway	2.94E-07	6.20E-05
hsa04144	Endocytosis	5.12E-07	1.10E-04
hsa04114	Oocyte meiosis	7.00E-07	1.50E-04
hsa05210	Colorectal cancer	7.63E-07	1.60E-04
hsa04660	T cell receptor signalling pathway	1.16E-06	2.40E-04
hsa05213	Endometrial cancer	2.74E-06	5.70E-04
hsa04914	Progesterone-mediated oocyte maturation	3.32E-06	6.80E-04
hsa04720	Long-term potentiation	3.78E-06	7.70E-04
hsa05160	Hepatitis C	1.04E-05	2.10E-03
hsa04810	Regulation of actin cytoskeleton	1.17E-05	2.40E-03
hsa04012	ErbB signalling pathway	1.38E-05	2.80E-03
hsa05216	Thyroid cancer	1.42E-05	2.80E-03
hsa04910	Insulin signalling pathway	1.45E-05	2.90E-03
hsa05211	Renal cell carcinoma	1.76E-05	3.50E-03
hsa03050	Proteasome	1.95E-05	3.80E-03
hsa05223	Non-small cell lung cancer	2.30E-05	4.50E-03
hsa03450	Non-homologous end-joining	5.12E-05	1.00E-02

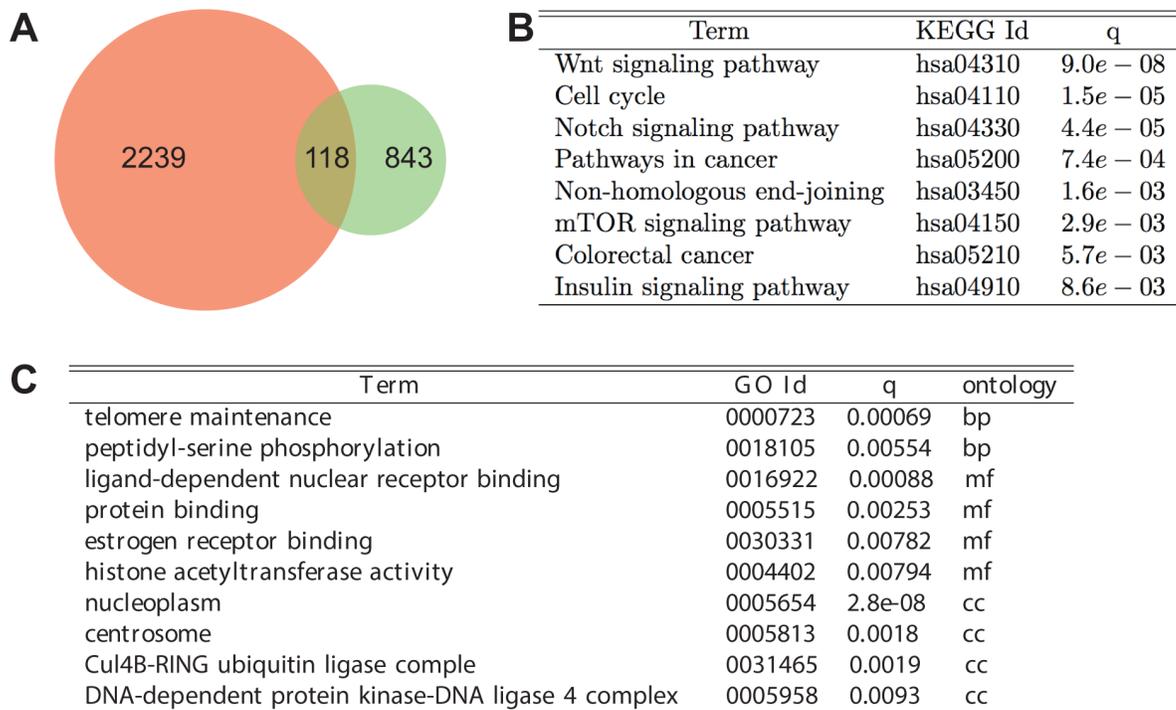
Only terms with  $q < 0.01$  (false discovery rate after Benjamini-Hochberg) are shown.

doi:10.1371/journal.pone.0126283.t002



**Fig 5. Homogenous functional spectrum of genes targeted by different ECCN genes.** The specific functions of genes interacting with each individual ECCN gene were counted and illustrated as heat map. Annotated KEGG pathways (A) and GO terms (B) found to be overrepresented amongst all predicted target genes in the preceding analysis were counted for each target gene, and these counts were then accumulated for each individual ECCN gene and represented as colours according to the legend. Rows and columns are ordered according to a hierarchical clustering.

doi:10.1371/journal.pone.0126283.g005



**Fig 6. Functional analysis of the consensus predicted ECCN target gene set.** Overlap between 2357 new ECCN elements (orange) based on expression pattern correlation and 961 genes obtained with text-mining methods (green) (A). This resulted in 118 new genes found by both methods. KEGG pathway annotation enrichment of the 118 consensus predicted genes (B) and corresponding GO enrichment (C).

doi:10.1371/journal.pone.0126283.g006

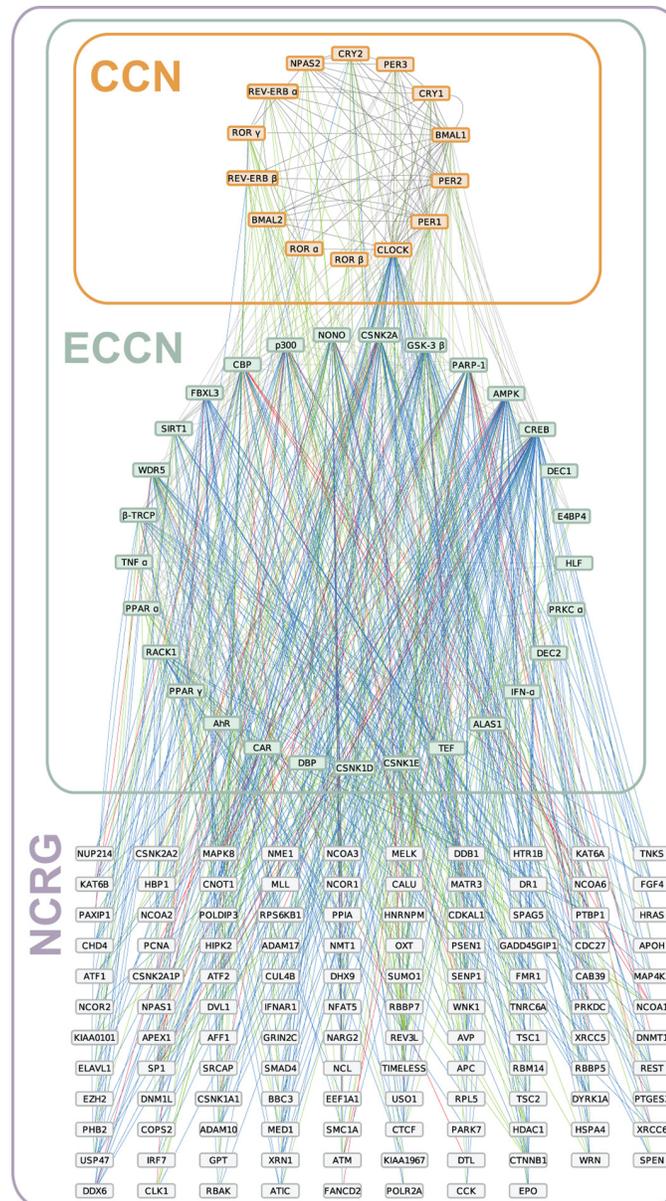
analysis of the 118 target genes using KEGG annotations indicated a strong connection to signalling- and cancer-related pathways (Fig 6B). The GO enrichment yielded the terms “telomere maintenance” and “peptidyl-serine phosphorylation” as significantly enriched biological processes (Fig 6C). The molecular function “ligand-dependent nuclear receptor binding” was also found to be significantly enriched ( $q < 0.0009$ ).

Finally, we used this intersection of the text-mining analysis and co-expression analysis to extend the ECCN, resulting in a novel network of circadian regulated genes (NCRG) comprising 161 genes all together (Fig 7). An additional 220 interactions between the ECCN and the new NCRG were found amongst the text-mining dataset and 402 interactions within the co-expression data. The number of correlation-based interactions is less informative because, as we have shown above it is not a precise method to infer network topology. Since this assessment was derived from a mixture of various tissue types, the NCRG can be expected to be an aggregation of different tissue-specific interactions.

### Circadian phenotype amongst predicted ECCN extension genes

We tested how many of the 118 novel ECCN targets were found to exhibit circadian expression patterns in circadian data sets [14, 27]. Integration of these two mouse datasets, and mapping to human genes via HomoloGene yielded a total of 1771 circadian transcripts. These included the following 19 out of our 118 predicted targets (for  $p < 0.009$ , for  $p < 0.05$  we find 59% genes out of the 118-set to be circadian): *Adam17*, *ApoH*, *Avp*, *Chd4*, *Clk1*, *Cops2*, *Ddx6*, *Dhx9*, *Dnm1l*, *Hnrnpm*, *Ifnar1*, *Map4k3*, *Ncl*, *Nmt1*, *Ncoa1*, *Psen1*, *Phb2*, *Smad4*, *Sumo1* (Table 3).

Additionally, we were interested in the possible consequences of perturbing the newly identified genes in the circadian phenotype and checked whether any of the 118 predicted



**Fig 7. Network representation of CCN/ECCN network together with the 118 predicted target genes (NCRG).** Boxes represent individual genes, which are connected by lines reflecting interactions that are known (grey), predicted by co-expression (blue), text-mining (green), or by both (red). The sub-networks are indicated by rectangles, the CCN (orange), the ECCN (green), and NCRG (purple).

doi:10.1371/journal.pone.0126283.g007

ECCN-interacting genes were found to cause perturbations on the circadian clock in available siRNA datasets [28, 29]. We found hits for different circadian phenotypes a) long-period phenotype: *Csnk1a1* (*casein kinase 1, A 1*), *Mapk8* (*mitogen-activated protein kinase 8*), *Ncl* (*nucleolin*); b) high-amplitude phenotype: *Ddb1* (*damage-specific DNA binding protein 1*); and c) short-period phenotype: *Cops2* (*COP9 signalosome subunit 2*). Among these, *Ncl* and *Cops2* also showed a circadian expression pattern. *Ncl* yielded a JTK q-value of 6.16e-06, a period of 24h, and a phase of 18.5. *Cops2* yields a p-value of 0.007, a period of 28h, and phase 2.5. These findings are summarized in Table 3.

Table 3. Properties of the consensus predicted ECCN target genes.

entrezID	Gene Symbol	Chip-Seq Target Genes							Circadian Phenotype				Pathological phenotype
		RevErb $\alpha$ [16, 17]	RevErb $\beta$ [16, 17]	RevErb $\alpha/\beta$ [17]	ROR $\alpha$ [22]	ROR $\gamma$ [22]	ROR $\gamma$ [22]	circadian expression	high-A	long-T	short-T	OMIM	
102	ADAM10		x										Reticulate acropigmentation of Kitamura, 615537 (3)~(Alzheimer disease 18, susceptibility to), 615590 (3)
328	APEX1	x											
350	APOH	x	x	x	x	x	x						
466	ATF1			x									
471	AT1C	x											AICA-ribosiduria due to AT1C deficiency, 608688 (3)
551	AVP							x					Diabetes insipidus, neurohypophyseal, 125700 (3)
813	CALU		x			x							
885	CCK	x											
996	CDC27				x								
1108	CHD4			x				x					
1195	CLK1							x					3MC syndrome 2, 265050 (3)
1386	ATF2												
1452	CSNK1A1	x	x	x	x	x					x		
1459	CSNK2A2				x								
1499	CTNINB1	x											Colorectal cancer, somatic, 114500 (3) ~Hepatocellular carcinoma, somatic, 114550 (3)~Mental retardation, autosomal dominant 19, 615075 (3)~Ovarian cancer, somatic, 167000 (3)~Pilomatricoma, somatic, 132600 (3)
1642	DDB1				x						x		
1656	DDX6				x			x					
1660	DHX9							x					
1855	DVL1				x								
1859	DYRK1A	x	x	x									Mental retardation, autosomal dominant 7, 614104 (3)
1915	EEF1A1	x	x			x							
1994	ELAVL1												
2177	FANCD2							x					Fanconi anemia, complementation group D2, 227646 (3)
2547	XRCC6		x										
2875	GPT	x	x	x									
2905	GRIN2C	x	x	x									
3308	HSPA4									x			
3454	IFNAR1		x	x	x	x					x		

(Continued)

Table 3. (Continued)

entrezID	Gene Symbol	Chip-Seq Target Genes						Circadian Phenotype				Pathological phenotype	
		RevErb $\alpha$ [16, 17]	RevErb $\beta$ [16, 17]	RevErb $\alpha/\beta$ [17]	ROR $\alpha$ [22]	ROR $\gamma$ [22]	ROR $\gamma$ [22]	circadian expression	high- A	long- T	short- T		OMIM
4089	SMAD4	x						x					Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome, 175050 (3) ~Myhre syndrome, 139210 (3) ~Pancreatic cancer, somatic, 260350 (3) ~Polyposis, juvenile intestinal, 174900 (3)
4297	MLL	x	x	x									Leukemia, myeloid/lymphoid or mixed-lineage (2)~Wiedemann-Steiner syndrome, 605130 (3)
4299	AFF1	x	x	x									
4670	HNRNPM							x					
4691	NCL					x						x	
4830	NIME1	x		x									Neuroblastoma, 256700 (3)
4836	NMT1	x	x										
5430	POLR2A					x							
5469	MED1			x									
5478	PPIA			x									
5599	MAPK8	x		x									
5663	PSEN1					x							Acne inversa, familial, 3, 613737 (3) ~Alzheimer disease, type 3, 607822 (3) ~Alzheimer disease, type 3, with spastic paraparesis and apraxia, 607822 (3) ~Alzheimer disease, type 3, with spastic paraparesis and unusual plaques, 607822 (3) ~Cardiomyopathy, dilated, 1U, 613694 (3) ~Dementia, frontotemporal, 600274 (3) ~Pick disease, 172700 (3)
5725	PTBP1	x		x									
5931	RBBP7												
5980	REV3L												
6125	RPL5			x									Diamond-Blackfan anemia 6, 612561 (3)
6667	SP1	x		x									
6868	ADAM17												Inflammatory skin and bowel disease, neonatal, 614328 (3)
7248	TSC1	x		x									Focal cortical dysplasia, Taylor balloon cell type, 607341 (3) ~Lymphangiomyomatosis, 606690 (3) ~Tuberous sclerosis-1, 191100 (3)
7341	SUMO1												Orofacial cleft 10, 613705 (3)
7520	XRCC5			x									
7994	KAT6A	x											
8021	NUP214			x									Leukemia, T-cell acute lymphoblastic (3) ~Leukemia, acute myeloid, 601626 (3)

(Continued)

Table 3. (Continued)

entrezID	Gene Symbol	ChIP-Seq Target Genes							Circadian Phenotype			Pathological phenotype	
		RevErb $\alpha$ [16, 17]	RevErb $\beta$ [16, 17]	RevErb $\alpha/\beta$ [17]	ROR $\alpha$ [22]	ROR $\gamma$ [22]	ROR $\gamma$ [22]	circadian expression	high-A	long-T	short-T		
8202	NCOA3	x	x	x									
8491	MAP4K3	x	x					x					
8615	USO1	x	x	x									
8648	NCOA1							x					
9318	COPS2											x	
9611	NCOR1	x	x	x									
9612	NCOR2	x	x	x									
10059	DNM1L								x				Encephalopathy, lethal, due to defective mitochondrial peroxisomal fission, 614388 (3)
10432	RBM14		x	x			x		x				
10499	NCOA2	x											
10615	SPAG5		x				x		x				
10664	CTCF							x					Mental retardation, autosomal dominant 21, 615502 (3)
10725	NFAT5							x					
10728	PTGES3									x			
11331	PHB2												
23013	SPEN	x											Megakaryoblastic leukemia, acute (2)
26959	HBP1	x											
27113	BBC3	x											
27327	TNRC6A												
28996	HIPK2	x	x										
51514	DTL	x	x	x									
54464	XRN1												
55031	USP47		x										
57786	RBAK									x			
79664	NARG2	x	x	x									
90480	GADD45GIP1												

Genes marked in bold were found to be BMAL1 targets.

All 62 genes (of 118) are shown which exhibit at least one of the following properties: regulated by REV-ERB or ROR, circadian expression pattern, causing a clock phenotype upon RNAi knockdown, predicted as similar to known clock genes [29], and featuring an OMIM annotation.

doi:10.1371/journal.pone.0126283.t003

We further compared our findings with a recent list of 1000 genes classified as—“sufficiently similar”—to known clock genes by a machine learning approach on a combination genome-scale datasets from mouse fibroblast cell lines [29]. One quarter of these genes were also contained in at least one of our gene sets (253 of 993 with a homolog in the human genome), and 10 genes were also detected by our text-mining and co-expression analysis: *Atf2*, *Ddx6*, *Dhx9*, *Elavl1*, *Hspa4*, *Ncl*, *Nme1*, *Med1*, *Rbbp7*, *Dnm1l*. Out of these, the four genes *Ddx6*, *Dhx9*, *Ncl*, and *Dnm1l* exhibit a circadian expression pattern.

### Clock target genes could be validated with ChIP-seq data

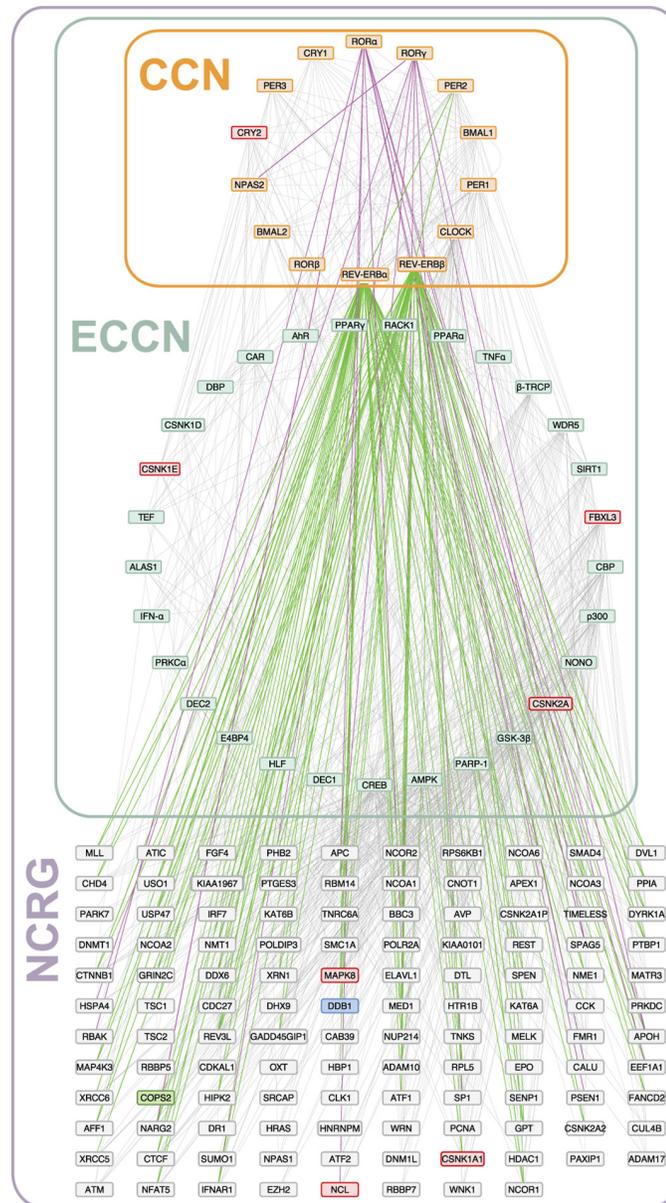
To further validate our 118 consensus genes gained from the bioinformatics approach, we examined the publicly available ChIP-seq datasets for REV-ERB $\alpha/\beta$  [16, 17]. Additionally, a BMAL1 dataset [12] was considered. ChIP-seq peak locations were used to calculate an association score (“ClosestGene” [30]) for each gene to the corresponding transcription factor. Simple threshold calculation then yielded a TF–target prediction. The gene association score  $S_{g,tf}$  was calculated for all annotated refSeq genes of the mouse genome build used in the corresponding experiment. The resulting log<sub>2</sub> transformed  $S_{g,tf}$  distributions are shown in S7 Fig. The threshold for accepting a TF–gene association was chosen as 3, which yields the higher second gene-score peak in case of the bimodal REV-ERB $\beta$  peak set, or the prominent right shoulder of the distribution for all other peak sets (S7 Fig). A total of 3847 predicted REV-ERB $\alpha$  and 3388 REV-ERB $\beta$  target genes [16] were found. The alternative dataset provided 4618 target genes associated with REV-ERB $\alpha/\beta$  unspecific peaks [17]. Lastly, this procedure yielded 223 significant BMAL1 target genes [12]. Since the ChIP-seq peak location data for ROR $\alpha$  and  $\gamma$ , were not accessible, we relied on the list of predicted targets provided by the authors based on a less stringent target prediction method [22, 31].

Overall, we obtained a set of 118 genes potentially regulated by the ECCN. Of those, 19 exhibited circadian expression patterns, 5 exhibited phenotypic changes in the clock when targeted with RNAi, 59 were targeted by REV-ERB $\alpha/\beta$ , and 14 were targeted by ROR $\alpha$  or  $\gamma$ . Additionally, the two NCRG genes *Ddb1* and *Mapk8* were found to associate with BMAL1 binding sites. These findings are summarized in Table 3 and depicted in Fig 8, (see S7 Table for all annotations).

## Discussion

The mammalian circadian clock is an endogenous, time-generating system with the peculiarity of synchronizing and propagating time-cues to the entire organism. Its relevance in the time-dependent regulation of biological processes has been shown at the organismal and cellular levels. As such, it is of no surprise that malfunctions of the circadian system were found to be associated to pathological phenotypes including obesity, sleep disorders and increasing incidence of cancer. The prospect of using individual patient-timing, based on the internal circadian clock, for therapy optimization is being explored with promising results. For instance, advances in chronotherapy have proven to be efficient in reducing toxicity and increasing efficacy in some types of cancer, particularly colon cancer [32]. A more detailed knowledge of the circadian network including the pathways it regulates is of major importance for the analysis on how time effects may be propagated and to determine the time-dependent action of certain drugs.

In this work we set up to dissect such clock-regulated pathways and to analyse the extent of circadian regulation at the cellular level by expanding the core circadian network to its potential target genes. We used human high-throughput transcriptome-data sets associated to text-mining of biomedical literature, for *de novo* clock regulated gene discovery.



**Fig 8. Transcriptional regulation of the CCN/ECCN extension network of 118 genes by REV-ERB and ROR.** Regulatory interactions of REV-ERB  $\alpha/\beta$  with NCRG genes (green lines) were derived from the locations of physical binding of these proteins in two ChIP-seq experiments [16, 17]. The ROR  $\alpha/\gamma$  interactions (purple lines) were adopted from the report of a third ChIP-seq experiment [22]. Genes with an observed phenotype in the genome-wide RNAi screen [28, 29] are shown with a coloured box, red indicating long period, blue a high amplitude, and green a short period.

doi:10.1371/journal.pone.0126283.g008

### A network of circadian regulation: combining independent evidences

Gene co-expression has previously been used to predict gene functions. Such works rely on the Pearson correlation coefficient and extensions of it and, although able to predict gene functions in mammals, are limited in terms of *de novo* network generation [24, 25]. We observed that reportedly interacting ECCN genes feature correlation values which are similar to non-reported. This is a limitation of co-expression methodologies and the problem of erroneous transitive

links inferred by correlation analysis was described before [33]. Therefore, we used a hybrid-methodology where to the expression correlation data we associated the text mining as an independent source of knowledge, enabling us to find regulated genes and their connection to the ECCN with increased confidence (Fig 1). This allowed us to partially overcome the limitations of expression analysis in terms of network topology and to be able to generate a semi-regulatory network for the mammalian circadian clock. Still, we do not analyse tissue-specificity issues which go beyond the scope of this work. Nevertheless, the circadian clock has been reported, in mammals, to be present in all cells so that the core network is expected to be very similar [34]. The output genes in the large network might indeed show tissue-specific differences which will be very interesting to explore in future work.

## Biological significance and impact in tumorigenesis

The detailed analysis of the network generated by our pipeline (NCRG) strengthens previous findings which associate the circadian clock to regulation of several molecular processes such as mRNA processing, cell division, cell cycle progression and DNA repair [19, 21, 35–40]. Particular pathways, including RNA transport, splicing and several cancer related pathways were identified by our study as being significantly associated with the circadian clock, highlighting the important function of the circadian system in the regulation of cellular processes. By comparing the difference in overrepresented terms between genes tightly- and those loosely-associated to the ECCN, we found that cell cycle and translation related terms are highly significant in loosely associated genes in comparison to tightly associated genes. We also found that splicing remains a highly over-represented term regardless of tightness (Fig 4). Together with the enriched biological processes such as “DNA-dependent regulation of transcription” and “gene expression”, it became clear that the co-expression based predicted ECCN target gene set has a stout emphasis on cellular signalling, transcriptional regulation, and cancer (Fig 5). Furthermore, several members of the predicted set of ECCN target genes are associated with Mendelian diseases as listed in the Online Mendelian Inheritance in Men dataset (OMIM) (S6 Table). 30% of the correlation/text-mining consensus genes featured such an annotation (35 of 118), pointing to the role of the circadian clock in pathogenesis.

In particular, among our top candidate genes is a group of genes associated with tumorigenesis (see Table 3): *Elavl1* is known to be highly expressed in several cancers and potentiates a characteristic pro-inflammatory profile of some immunological and non-immunological diseases [41], *Nme1* is considered a tumour suppressor and its expression is reduced in metastatic cancers [42], *Dhx6* belongs to the DEAD box helicase superfamily and is involved in DNA repair, *Med1* regulates p53-dependent apoptosis [43] and *Rbbp7* interacts with the tumour-suppressor gene *Brca1* [44] and may have a role in the regulation of cell proliferation and differentiation.

Remarkably, we found a subset of nine genes (*ApoH*, *Ifnar1*, *Sp1*, *Narg2*, *CALU*, *EEF1A1*, *RBM14*, *Spag5*, *Med1*) which are targets of both REV-ERB and ROR according to Chip-Seq experiments. These two nuclear receptors are known to bind RORE elements within the promoter regions of target genes: while REV-ERB is an inhibitor, ROR acts as an activator. APOH (Apolipoprotein H) and IFNAR1 (Interferon Alpha, Beta, Omega Receptor) are involved in immune disorders [45, 46]. SP1 (Sp1 transcription factor) is also involved in immune response and in many other cellular processes, including cell differentiation, cell growth, apoptosis, response to DNA damage, and chromatin remodelling [19]. NARG2 (NMDA receptor regulated 2) is associated to breast cancer [47], and *Med1* regulates p53-dependent apoptosis [43] and was found to be mutated in human carcinomas with microsatellite instability [48]. The eukaryotic translation elongation factor EEF1A1 was recently shown to mediate the alternative

caspace-independent cell death mechanism induced by genetically unstable tetraploidy [49]. The sperm associated antigen 5 (SPAG5) was found to be associated with various types of cancer, such as cervical cancer and breast cancer [50]. Circadian regulation of these genes and as such of the processes they regulate could be achieved via a fine-tuning of ROR/REV-ERB.

Two other circadian regulated genes identified by our study are *nucleolin* (*Ncl*) and *Ddx6*. The analysis of ChIP-seq data identified these genes as targets of ROR $\gamma$  and REV-ERB $\alpha$ ,  $\beta$ , respectively. Interestingly, they were also reported to be involved in miRNA regulation [51–53]. DDX6 (RNA helicase) is found in p-bodies for mRNA degradation, needed for miRNA-mediated silencing. NCL regulates several miRNAs including miR-21, miR-221, miR222 and miR-103. miR-21 is defined as an oncogene and found to be overexpressed in most tumour types [51, 54–59], whereas miR-221 and miR222 show an increased expression in human breast cancer [60, 61]. Also, miR-222 was shown to promote resistance of cancer cells to cytotoxic T lymphocytes [62]. Interestingly, miR-103 which is also a target of NCL was reported to exhibit circadian pattern [63].

Altogether, our data allowed the generation of a large network of circadian regulation. The network was retrieved from human expression data intersected with text-mining of the biomedical literature, for topology refinement and *de novo* target identification. The novel predicted targets of the circadian clock network showed a remarkable association to cancer driving mechanisms. One of these mechanisms is miRNA regulation. Very recent studies point to an influence of miRNAs on the circadian clock [64–71], but only a few links on the regulation of miRNAs via the circadian clock have been described [69]. NCL represents a potential novel link via which the circadian clock, in particular ROR $\gamma$ , regulates the expression of miRNAs, with particular consequences in cancer progression.

## Methods

### Preprocessing

For all text-mining steps we used articles from PubMed and PubMed Central open access subset.

### Named entity recognition

Genes: For gene name recognition and normalization we used the GNAT library [72]. GNAT uses custom dictionaries and conditional random fields (CRF) for gene name recognition and subsequently normalises gene mentions to Entrez Gene ID's. The system is ranked among the first in several critical evaluations [73, 74] and achieves, according to these assessments, a precision of 82% and recall of 82% for abstracts and 54/47% for full—text articles.

### Relation extraction

GeneView (a search engine which uses a comprehensively annotated database of all PubMed abstracts and 270,000 full texts from the open PubMed Central corpus) uses the shallow linguistic kernel [75] and LibSVM for relationship extraction between proteins. The model is trained on the ensemble of five publicly available training corpora [76]. This kernel achieved very good results in a comprehensive evaluation of nine machine learning kernels for PPI extraction from text [77–79]. Furthermore, it does not use dependency information and thus is very fast, a pre-requisite for usage in a large system such as GeneView. Data contained in GeneView is available at <http://bc3.informatik.hu-berlin.de/>. To account for species specificity, we mapped mammalian gene identifiers to Homologene clusters [80]. To test the efficiency of text-mining in contributing to new network generation, we first evaluated its ability to

reconstruct a previously designed network of clock-controlled genes (CCGs) containing 121 interactions among 41 different proteins [19]. We used GeneView to extract all pairwise interactions. GeneView contained evidence for 73% of all interactions described in the network tested. The high sensitivity of the method encouraged us to further develop our pipeline in order to ascertain potential new elements and interactions. We further used GeneView to collect all interactions among the CCN and its directly interacting neighbours. After curation and filtering for direct interactions, we enriched the core-clock network with 108 novel interactions supported by 132 PubMed references, which led to the extended core-clock network (ECCN) recently reported [21]. For the ECCN, each candidate interaction is supported by up to 851 sentences (in total 4,206 sentences). We reduced the number of sentences to 580 by ranking them by confidence and returning only 5 sentences at maximum for each candidate. Sentences containing potentially novel PPI were ranked by the confidence of the classifier (ie. distance to the hyperplane) and were subsequently evaluated.

### Predicting interactions using coexpression data and overrepresentation of associated gene terms

Each dataset was assessed on the number of genes they share with the ECCN and how well the correlation coefficient distributions of known ECCN gene interactions were separated from a background distribution of all genes, where the Wilcoxon Rank Sum test was used for quantification. For more details on the dataset properties and selection, see [S2 Text](#).

To find associated genes based on the correlation coefficients, we selected the 10000 highest correlations between any ECCN gene and a non-ECCN gene as predicted interactions, thereby considering the 1.18% most extreme correlation values.

We sought to detect and characterize only genes that were tightly associated with the ECCN, where "tightness" was defined as the number of connections between a gene and a set of genes. Accordingly, the comparison of the number of predicted NCRG with required tightness 1 to 10 shows the most drastic decline between 1 and 2, which quickly diminishes with rising tightness values ([Fig 4](#)). We therefore chose to employ a tightness threshold of 2 for the remaining analysis. We then proceeded to find the overrepresented terms and enriched clusters using the R package TopGO [81]. We annotated the associated genes with terms from the Genetic Association Database [82], Online Mendelian Inheritance in Man database [83], Swissprot Protein Information Resource [84], Gene Ontology [85], Pubmed and Kyoto Encyclopedia of Genes and Genomes [86]. Significant overrepresentation was determined using p-values corrected by Benjamini-Hochberg multiple testing correction (q-values).

### Integration of the predicted NCRG with transcriptional features

We compared our NCRG prediction with the machine learning based prediction of clock genes [29]. Therefore, we retrieved the top 1000 genes as of the evidence factor ranks and used the HomoloGene database build 66 [80] to map the reported mouse genes to 993 unique entrez genes, could then be compared to our predicted genes set.

Similarly, we tested how many of the NCRG are amongst the genes with circadian expression regulation according to recent publications [14, 27]. After combination of both lists of mouse genes, a total of 1771 unique entrez transcripts were obtained for comparison after mapping via HomoloGene build 68.

An extensive collection of genes which lead to circadian clock phenotypes upon knockout via RNAi has been described recently [28]. The reported 343 genes are categorized into double hitters, i.e. two different pairs of siRNAs lead to a circadian clock phenotype, and single hitters,

for which only one of the two siRNA pairs designed for each gene lead to a phenotype, where amplitude- and phase-changes were considered as phenotype.

## ChIP-seq data analysis

We employed the R package TFFTargetCaller [30] to derive target gene sets for clock-related transcription factors from experimental Chip-seq data using the method “ClosestGene”. We used available data sets to extract target genes for REV-ERB  $\alpha/\beta$  [16, 17] and for BMAL1 [12]. These include all available Chip-seq data sets for core-clock genes. Specifically, the genomic peak locations were obtained, and the gene association score  $S_{g,tf}$  was calculated for all annotated refSeq genes of the mouse genome build used in the corresponding experiment. The resulting  $\log_2$  transformed  $S_{g,tf}$  distributions are shown in S7 Fig. The threshold for accepting a TF–gene association was chosen as 3, which yields the higher second gene-score peak in case of the bimodal REV-ERB  $\beta$  peak set (S7B Fig), or the prominent right shoulder of the distribution for all other peak sets. Since the genomic locations of the peaks for the ROR  $\alpha/\gamma$  dataset were not available, we used the predicted target list provided by the authors [22, 31].

## Supporting Information

### S1 Fig. Correlation distributions for clock network gene pairs versus random gene pairs.

Cumulative correlation value distributions obtained from the HSA dataset, shown for comparison with Fig 3 in the main text.

(EPS)

**S2 Fig. Correlation of reported CCN interactions, ECCN interactions as compared to random background.** The Pearson  $\rho$  distributions of pairs of ECCN genes reported to interact but excluding CCN (ECCN, green), reported pairs of CCN genes (CCN, orange), and 43 randomly chosen genes versus all genes as background (BG, black) are shown. Pearson correlation coefficient (A, C) and mutual rank (B, D) probability density functions for known CCN interactions and reported ECCN interactions compared with random background. Data extracted from the HSA and HSA2 dataset are shown in (A, B), and (C, D) respectively. Shown for comparison with Fig 3 in the main text.

(EPS)

### S3 Fig. Correlation of reported ECCN interactions compared to not-reported interactions.

Pearson correlation coefficient (A, C) and mutual rank (B, D) probability density functions for all ECCN interactions (i.e. including all CCN interactions, “reported ECCN”) compared with all other possible pairs between ECCN genes, for which no interaction is reported (“other”). Data extracted from the Hsa and Hsa2 dataset are shown in (A, B), and (C, D), respectively. Shown for comparison with Fig 3 in the main text.

(EPS)

**S4 Fig. Correlation of the ECCN interactions and the background.** Pearson correlation coefficient (A, C) and mutual rank (B, D) probability density functions for all possible pairs of ECCN genes compared with all other possible pairs between one of the 43 ECCN genes and a non-ECCN gene as background. Data extracted from the Hsa and Hsa2 dataset are shown in (A, B), and (C, D), respectively. Shown for comparison with Fig 3 in the main text.

(EPS)

**S5 Fig. Tightness filtering effect.** The number of predicted target genes (y-axis) decreases when increasing the minimal number of ECCN genes, with which it has to correlate (x-axis).

(EPS)

**S6 Fig. Number of interactions (x-axis) for 32 ECCN genes (y-axis) as predicted by expression correlation.** The remaining 11 ECCN genes do not feature predicted targets.  
(EPS)

**S7 Fig. Association scores for core clock transcription factors.** Association strength scores  $S_{tf,g}$  between the core clock transcription factors REV-ERB  $\alpha/\beta$  and BMAL1 and all refSeq genes annotated in the corresponding genome version (mm8 [17] or mm9 otherwise) were calculated using the “ClosestGene” method of the R package TFTargetCaller and the ChIP-seq peak annotations. The number of genes with  $S_{tf,g} > 0$  is shown as  $n_{gene}$ , the cutoff for accepted TF-gene association was set to 3 as marked with a red dashed line, and the number of accepted target genes is shown as  $n_{target}$  for each individual dataset.  
(EPS)

**S8 Fig. Cytoscape file for Fig 7.** (<http://www.cytoscape.org/>).  
(ZIP)

**S9 Fig. Cytoscape file for Fig 8.** (<http://www.cytoscape.org/>).  
(ZIP)

**S1 Table. List of ECCN interactions with publication references obtained by text-mining.**  
(XLS)

**S2 Table. Extension of the ECCN network using text-mining method.**  
(XLSX)

**S3 Table. List of GO term annotations enriched amongst the ECCN target genes predicted by text-mining.** The table provides the term ids (“GO.ID”), the corresponding “Term”, as well as the overrepresentation p-value after false discovery rate correction after Benjamini-Hochberg (“pval”). In addition, the total number of genes annotated with the respective term is provided (“Annotated”), the number of significant annotations (“Significant”), along with the number of annotations expected by chance in the gene set (“Expected”).  
(XLSX)

**S4 Table. List of KEGG pathway annotations enriched amongst the ECCN target genes predicted by text-mining.** The table provides the pathway ids (“kegg.id”), the corresponding pathway “name”, as well as the overrepresentation p-value before (“pval”) and after false discovery rate correction after Benjamini-Hochberg (“fdr”).  
(XLSX)

**S5 Table. List of all newly predicted interactions.** The columns “gene1.entrez”, “gene2.entrez”, “gene1.symbol”, and “gene2.symbol” provide the Entrez ids and gene symbols for the two predicted interacting genes, respectively. The Boolean flags “txtmn” and “coxp” indicate interactions predicted by text-mining, and co-expression, respectively. “with.consensus” indicates interactions involving one of the 118 consensus genes, and “overlapping” indicates the interactions predicted similarly by text-mining and co-expression.  
(XLSX)

**S6 Table. Characterization of the 118 predicted NCRG regarding disease-related annotation.** Annotations from OMIM, gad, and the KEGG database were integrated in addition to SNPs. Genetic association database (gad), KEGG pathway annotation, traits that significantly associate with the gene as provided in the Ensemble database (gwas\_trait, gwas\_pval, gwas\_-pubmed\_id), the number of non-synonymous SNPs found in the gene (nonsyn\_count, nonsyn\_norm), the Uniprot database derived protein domains (up\_seq\_feature), the Gene

Ontology biological processes annotations (goterm\_bp), and the Gene Reference into Function (generif).  
(XLS)

**S7 Table. Characterization of the 118 NCRG regarding expression regulation by transcription factors.** The transcription factor target prediction of all 118 NCRG for both REV-ERB $\alpha/\beta$  datasets, the BMAL1, and also the ROR $\alpha/\gamma$  dataset is provided. Additionally, the phenotype upon gene knockdown observed [28] and the prediction of “similar to clock gene” [29] are included. The observed circadian expression and OMIM annotations are indicated.  
(XLS)

**S1 Text. Text-mining-based assembly and characterization of the ECCN network.**  
(DOCX)

**S2 Text. Comparison of available co-expression databases.**  
(DOCX)

## Acknowledgments

We thank members of the Relógio group for critical comments and technical support.

## Author Contributions

Conceived and designed the experiments: AR RL. Performed the experiments: LC PT RL. Analyzed the data: AR LF MA HH PT RL UL. Contributed reagents/materials/analysis tools: AR LC PT RL. Wrote the paper: AR RL. Critically read and contributed for writing the manuscript: AR LF MA HH PT RL UL.

## References

1. Lowrey PL, Takahashi JS. Genetics of circadian rhythms in Mammalian model organisms. *Advances in genetics*. 2011; 74:175–230. doi: [10.1016/B978-0-12-387690-4.00006-4](https://doi.org/10.1016/B978-0-12-387690-4.00006-4) PMID: [21924978](https://pubmed.ncbi.nlm.nih.gov/21924978/).
2. Albrecht U. Timing to perfection: the biology of central and peripheral circadian clocks. *Neuron*. 2012; 74:246–60. doi: [10.1016/j.neuron.2012.04.006](https://doi.org/10.1016/j.neuron.2012.04.006) PMID: [22542179](https://pubmed.ncbi.nlm.nih.gov/22542179/).
3. Bass J. Circadian topology of metabolism. *Nature*. 2012; 491:348–56. doi: [10.1038/nature11704](https://doi.org/10.1038/nature11704) PMID: [23151577](https://pubmed.ncbi.nlm.nih.gov/23151577/).
4. Kondratova AA, Kondratov RV. The circadian clock and pathology of the ageing brain. *Nature reviews Neuroscience*. 2012; 13:325–35. doi: [10.1038/nrn3208](https://doi.org/10.1038/nrn3208) PMID: [22395806](https://pubmed.ncbi.nlm.nih.gov/22395806/).
5. Takahashi JS, Hong HK, Ko CH, McDearmon EL. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nature reviews Genetics*. 2008; 9:764–75. doi: [10.1038/nrg2430](https://doi.org/10.1038/nrg2430) PMID: [18802415](https://pubmed.ncbi.nlm.nih.gov/18802415/).
6. Levi F, Schibler U. Circadian rhythms: mechanisms and therapeutic implications. *Annual review of pharmacology and toxicology*. 2007; 47:593–628. doi: [10.1146/annurev.pharmtox.47.120505.105208](https://doi.org/10.1146/annurev.pharmtox.47.120505.105208) PMID: [17209800](https://pubmed.ncbi.nlm.nih.gov/17209800/).
7. Saini C, Suter DM, Liani A, Gos P, Schibler U. The mammalian circadian timing system: synchronization of peripheral clocks. *Cold Spring Harbor symposia on quantitative biology*. 2011; 76:39–47. doi: [10.1101/sqb.2011.76.010918](https://doi.org/10.1101/sqb.2011.76.010918) PMID: [22179985](https://pubmed.ncbi.nlm.nih.gov/22179985/).
8. Bollinger T, Schibler U. Circadian rhythms—from genes to physiology and disease. *Swiss medical weekly*. 2014; 144:w13984. doi: [10.4414/smw.2014.13984](https://doi.org/10.4414/smw.2014.13984) PMID: [25058693](https://pubmed.ncbi.nlm.nih.gov/25058693/).
9. Relogio A, Westermark PO, Wallach T, Schellenberg K, Kramer A, Herzel H. Tuning the mammalian circadian clock: robust synergy of two loops. *PLoS computational biology*. 2011; 7(12):e1002309. doi: [10.1371/journal.pcbi.1002309](https://doi.org/10.1371/journal.pcbi.1002309) PMID: [22194677](https://pubmed.ncbi.nlm.nih.gov/22194677/)
10. Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, et al. System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nature genetics*. 2005; 37:187–92. doi: [10.1038/ng1504](https://doi.org/10.1038/ng1504) PMID: [15665827](https://pubmed.ncbi.nlm.nih.gov/15665827/).

11. Zhang EE, Kay SA. Clocks not winding down: unravelling circadian networks. *Nature reviews Molecular cell biology*. 2010; 11:764–76. doi: [10.1038/nrm2995](https://doi.org/10.1038/nrm2995) PMID: [20966970](https://pubmed.ncbi.nlm.nih.gov/20966970/).
12. Rey G, Cesbron FC, Rougemont J, Reinke H, Brunner M, Naef F. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS biology*. 2011; 9(2):e1000595. doi: [10.1371/journal.pbio.1000595](https://doi.org/10.1371/journal.pbio.1000595) PMID: [21364973](https://pubmed.ncbi.nlm.nih.gov/21364973/)
13. Koike N, Yoo S-H, Huang H-C, Kumar V, Lee C, Kim T-K, et al. Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science (New York, NY)*. 2012.
14. Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, et al. Harmonics of circadian gene transcription in mammals. *PLoS genetics*. 2009; 5(4):e1000442. doi: [10.1371/journal.pgen.1000442](https://doi.org/10.1371/journal.pgen.1000442) PMID: [19343201](https://pubmed.ncbi.nlm.nih.gov/19343201/)
15. Feng D, Liu T, Sun Z, Bugge A, Mullican SE, Alenghat T, et al. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science (New York, NY)*. 2011; 331(6022):1315–9. doi: [10.1126/science.1198125](https://doi.org/10.1126/science.1198125) PMID: [21393543](https://pubmed.ncbi.nlm.nih.gov/21393543/)
16. Cho H, Zhao X, Hatori M, Yu RT, Barish GD, Lam MT, et al. Regulation of circadian behaviour and metabolism by Rev-Erb- $\alpha$  and Rev-Erb- $\beta$ . *Nature*. 2012; 485(7396):123–7. Epub 2012/03/31. doi: [10.1038/nature11048](https://doi.org/10.1038/nature11048) PMID: [22460952](https://pubmed.ncbi.nlm.nih.gov/22460952/); PubMed Central PMCID: [PMCPmc3367514](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC3367514/).
17. Bugge A, Feng D, Everett LJ, Briggs ER, Mullican SE, Wang F, et al. Rev-erb  $\alpha$  and Rev-erb  $\beta$  coordinately protect the circadian clock and normal metabolic function. *Genes & development*. 2012; 26(7):657–67.
18. Korencic A, Kosir R, Bordyugov G, Lehmann R, Rozman D, Herzel H. Timing of circadian genes in mammalian tissues. *Scientific reports*. 2014; 4:5782. doi: [10.1038/srep05782](https://doi.org/10.1038/srep05782) PMID: [25048020](https://pubmed.ncbi.nlm.nih.gov/25048020/).
19. Bozek K, Relogio A, Kielbasa SM, Heine M, Dame C, Kramer A, et al. Regulation of clock-controlled genes in mammals. *PLoS One*. 2009; 4(3):e4882. doi: [10.1371/journal.pone.0004882](https://doi.org/10.1371/journal.pone.0004882) PMID: [19287494](https://pubmed.ncbi.nlm.nih.gov/19287494/); PubMed Central PMCID: [PMC2654074](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC2654074/).
20. Wallach T, Schellenberg K, Maier B, Kalathur RKR, Porras P, Wanker EE, et al. Dynamic circadian protein-protein interaction networks predict temporal organization of cellular functions. *PLoS genetics*. 2013; 9(3):e1003398. doi: [10.1371/journal.pgen.1003398](https://doi.org/10.1371/journal.pgen.1003398) PMID: [23555304](https://pubmed.ncbi.nlm.nih.gov/23555304/)
21. Relogio A, Thomas P, Medina-Perez P, Reischl S, Bervoets S, Gloc E, et al. Ras-mediated deregulation of the circadian clock in cancer. *PLoS genetics*. 2014; 10(5):e1004338. doi: [10.1371/journal.pgen.1004338](https://doi.org/10.1371/journal.pgen.1004338) PMID: [24875049](https://pubmed.ncbi.nlm.nih.gov/24875049/)
22. Takeda Y, Kang HS, Freudenberg J, DeGraff LM, Jothi R, Jetten AM. Retinoic acid-related orphan receptor  $\gamma$  (ROR $\gamma$ ): a novel participant in the diurnal regulation of hepatic gluconeogenesis and insulin sensitivity. *PLoS genetics*. 2014; 10(5):e1004331. doi: [10.1371/journal.pgen.1004331](https://doi.org/10.1371/journal.pgen.1004331) PMID: [24831725](https://pubmed.ncbi.nlm.nih.gov/24831725/)
23. Thomas P, Starlinger J, Vowinkel A, Arzt S, Leser U. GeneView: a comprehensive semantic search engine for PubMed. *Nucleic acids research*. 2012; 40(Web Server issue):W585–91. doi: [10.1093/nar/gks563](https://doi.org/10.1093/nar/gks563) PMID: [22693219](https://pubmed.ncbi.nlm.nih.gov/22693219/); PubMed Central PMCID: [PMC3394277](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC3394277/).
24. Nayak RR, Kearns M, Spielman RS, Cheung VG. Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome research*. 2009; 19(11):1953–62. doi: [10.1101/gr.097600.109](https://doi.org/10.1101/gr.097600.109) PMID: [19797678](https://pubmed.ncbi.nlm.nih.gov/19797678/)
25. Obayashi T, Hayashi S, Shibaoka M, Saeki M, Ohta H, Kinoshita K. COXPRESdb: a database of coexpressed gene networks in mammals. *Nucleic acids research*. 2008; 36(Database issue):D77–82. PMID: [17932064](https://pubmed.ncbi.nlm.nih.gov/17932064/)
26. Prieto C, Risueño A, Fontanillo C, De las Rivas J. Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PLoS one*. 2008; 3:e3911. doi: [10.1371/journal.pone.0003911](https://doi.org/10.1371/journal.pone.0003911) PMID: [19081792](https://pubmed.ncbi.nlm.nih.gov/19081792/).
27. Yan J, Wang H, Liu Y, Shao C. Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLoS Comput Biol*. 2008; 4(10):e1000193. Epub 2008/10/11. doi: [10.1371/journal.pcbi.1000193](https://doi.org/10.1371/journal.pcbi.1000193) PMID: [18846204](https://pubmed.ncbi.nlm.nih.gov/18846204/); PubMed Central PMCID: [PMCPmc2543109](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC2543109/).
28. Zhang EE, Liu AC, Hirota T, Miraglia LJ, Welch G, Pongsawakul PY, et al. A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell*. 2009; 139(1):199–210. doi: [10.1016/j.cell.2009.08.031](https://doi.org/10.1016/j.cell.2009.08.031) PMID: [19765810](https://pubmed.ncbi.nlm.nih.gov/19765810/)
29. Anafi RC, Lee Y, Sato TK, Venkataraman A, Ramanathan C, Kavakli IH, et al. Machine learning helps identify CHRONO as a circadian clock component. *PLoS biology*. 2014; 12(4):e1001840. doi: [10.1371/journal.pbio.1001840](https://doi.org/10.1371/journal.pbio.1001840) PMID: [24737000](https://pubmed.ncbi.nlm.nih.gov/24737000/)
30. Sikora-Wohlfeld W, Ackermann M, Christodoulou EG, Singaravelu K, Beyer A. Assessing Computational Methods for Transcription Factor Target Gene Identification Based on CHIP-seq Data. *PLoS Computational Biology*. 2013; 9(11):e1003342. doi: [10.1371/journal.pcbi.1003342](https://doi.org/10.1371/journal.pcbi.1003342) PMID: [24278002](https://pubmed.ncbi.nlm.nih.gov/24278002/)

31. Takeda Y, Jothi R, Birault V, Jetten AM. ROR $\gamma$  directly regulates the circadian expression of clock genes and downstream targets in vivo. *Nucleic acids research*. 2012; 40(17):8519–35. PMID: [22753030](#)
32. Lévi F, Focan C, Karaboué A, de la Valette V, Focan-Henrard D, Baron B, et al. Implications of circadian clocks for the rhythmic delivery of cancer therapeutics. *Advanced drug delivery reviews*. 2007; 59:1015–35. doi: [10.1016/j.addr.2006.11.001](#) PMID: [17692427](#).
33. Wright S. Correlation and causation. *J Agricultural Research*. 1921; 20:557–85.
34. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(45):16219–24. doi: [10.1073/pnas.1408886111](#) PMID: [25349387](#); PubMed Central PMCID: PMC4234565.
35. Padmanabhan K, Robles MS, Westerling T, Weitz CJ. Feedback regulation of transcriptional termination by the mammalian circadian clock PERIOD complex. *Science (New York, NY)*. 2012; 337:599–602. doi: [10.1126/science.1221592](#) PMID: [22767893](#).
36. Kang T-H, Reardon JT, Sancar A. Regulation of nucleotide excision repair activity by transcriptional and post-transcriptional control of the XPA protein. *Nucleic acids research*. 2011; 39:3176–87. doi: [10.1093/nar/gkq1318](#) PMID: [21193487](#).
37. Lévi F, Filipinski E, Iurisci I, Li XM, Innominato P. Cross-talks between circadian timing system and cell division cycle determine cancer biology and therapeutics. *Cold Spring Harbor symposia on quantitative biology*. 2007; 72:465–75. doi: [10.1101/sqb.2007.72.030](#) PMID: [18419306](#).
38. Unsal-Kaçmaz K, Mullen TE, Kaufmann WK, Sancar A. Coupling of human circadian and cell cycles by the timeless protein. *Molecular and cellular biology*. 2005; 25:3109–16. doi: [10.1128/MCB.25.8.3109-3116.2005](#) PMID: [15798197](#).
39. Bieler J, Cannavo R, Gustafson K, Gobet C, Gatfield D, Naef F. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular systems biology*. 2014; 10(7):739. doi: [10.15252/msb.20145218](#) PMID: [25028488](#).
40. Feillet C, Krusche P, Tamanini F, Janssens RC, Downey MJ, Martin P, et al. Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(27):9828–33. doi: [10.1073/pnas.1320474111](#) PMID: [24958884](#); PubMed Central PMCID: PMC4103330.
41. Srikantan S, Gorospe M. HuR function in disease. *Frontiers in bioscience (Landmark edition)*. 2012; 17:189–205. PMID: [22201738](#).
42. Fan Z, Beresford PJ, Oh DY, Zhang D, Lieberman J. Tumor Suppressor NM23-H1 Is a Granzyme A-Activated DNase during CTL-Mediated Apoptosis, and the Nucleosome Assembly Protein SET Is Its Inhibitor. *Cell*. 2003; 112:659–72. doi: [10.1016/S0092-8674\(03\)00150-8](#) PMID: [12628186](#)
43. Frade R, Balbo M, Barel M. RB18A regulates p53-dependent apoptosis. *Oncogene*. 2002; 21:861–6. doi: [10.1038/sj.onc.1205177](#) PMID: [11840331](#).
44. Chen GC, Guan LS, Yu JH, Li GC, Choi Kim HR, Wang ZY. Rb-associated protein 46 (RbAp46) inhibits transcriptional transactivation mediated by BRCA1. *Biochemical and biophysical research communications*. 2001; 284:507–14. doi: [10.1006/bbrc.2001.5003](#) PMID: [11394910](#).
45. Higashimoto M, Homma Y, Umetsu M, Konno Y, Ono K, Yoshimoto N, et al. Circadian rhythm of apoprotein H (beta2-glycoprotein-1) in human plasma. *Biochemical and biophysical research communications*. 2007; 360(2):418–22. doi: [10.1016/j.bbrc.2007.06.061](#) PMID: [17603016](#).
46. Takane H, Ohdo S, Yamada T, Koyanagi S, Yukawa E, Higuchi S. Relationship between diurnal rhythm of cell cycle and interferon receptor expression in implanted-tumor cells. *Life sciences*. 2001; 68(12):1449–55. PMID: [11388696](#).
47. Yamamoto T, Mencarelli MA, Di Marco C, Mucciolo M, Vascotto M, Balestri P, et al. Overlapping microdeletions involving 15q22.2 narrow the critical region for intellectual disability to NARG2 and RORA. *European journal of medical genetics*. 2014; 57(4):163–8. doi: [10.1016/j.ejmg.2014.02.001](#) PMID: [24525055](#).
48. Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, et al. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nature genetics*. 1999; 23:266–8. doi: [10.1038/15443](#) PMID: [10545939](#).
49. Kobayashi Y, Yonehara S. Novel cell death by downregulation of eEF1A1 expression in tetraploids. *Cell death and differentiation*. 2009; 16:139–50. doi: [10.1038/cdd.2008.136](#) PMID: [18820646](#).
50. Yuan L-J, Li J-D, Zhang L, Wang JH, Wan T, Zhou Y, et al. SPAG5 upregulation predicts poor prognosis in cervical cancer patients and alters sensitivity to taxol treatment via the mTOR signaling pathway. *Cell death & disease*. 2014; 5:e1247. doi: [10.1038/cddis.2014.222](#) PMID: [24853425](#).

51. Pichiorri F, Palmieri D, De Luca L, Consiglio J, You J, Rocci A, et al. In vivo NCL targeting affects breast cancer aggressiveness through miRNA regulation. *The Journal of experimental medicine*. 2013; 210(5):951–68. doi: [10.1084/jem.20120950](https://doi.org/10.1084/jem.20120950) PMID: [23610125](https://pubmed.ncbi.nlm.nih.gov/23610125/); PubMed Central PMCID: PMC3646490.
52. Yu JM, Wu X, Gimble JM, Guan X, Freitas MA, Bunnell BA. Age-related changes in mesenchymal stem cells derived from rhesus macaque bone marrow. *Aging cell*. 2011; 10(1):66–79. doi: [10.1111/j.1474-9726.2010.00646.x](https://doi.org/10.1111/j.1474-9726.2010.00646.x) PMID: [20969724](https://pubmed.ncbi.nlm.nih.gov/20969724/).
53. Iio A, Takagi T, Miki K, Naoe T, Nakayama A, Akao Y. DDX6 post-transcriptionally down-regulates miR-143/145 expression through host gene NCR143/145 in cancer cells. *Biochimica et biophysica acta*. 2013; 1829(10):1102–10. doi: [10.1016/j.bbagr.2013.07.010](https://doi.org/10.1016/j.bbagr.2013.07.010) PMID: [23932921](https://pubmed.ncbi.nlm.nih.gov/23932921/).
54. Medina PP, Nolde M, Slack FJ. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature*. 2010; 467(7311):86–90. doi: [10.1038/nature09284](https://doi.org/10.1038/nature09284) PMID: [20693987](https://pubmed.ncbi.nlm.nih.gov/20693987/).
55. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer research*. 2005; 65(14):6029–33. doi: [10.1158/0008-5472.CAN-05-0137](https://doi.org/10.1158/0008-5472.CAN-05-0137) PMID: [16024602](https://pubmed.ncbi.nlm.nih.gov/16024602/).
56. Ciafre SA, Galardi S, Mangiola A, Ferracin M, Liu CG, Sabatino G, et al. Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochemical and biophysical research communications*. 2005; 334(4):1351–8. doi: [10.1016/j.bbrc.2005.07.030](https://doi.org/10.1016/j.bbrc.2005.07.030) PMID: [16039986](https://pubmed.ncbi.nlm.nih.gov/16039986/).
57. Hashimi ST, Fulcher JA, Chang MH, Gov L, Wang S, Lee B. MicroRNA profiling identifies miR-34a and miR-21 and their target genes JAG1 and WNT1 in the coordinate regulation of dendritic cell differentiation. *Blood*. 2009; 114(2):404–14. doi: [10.1182/blood-2008-09-179150](https://doi.org/10.1182/blood-2008-09-179150) PMID: [19398721](https://pubmed.ncbi.nlm.nih.gov/19398721/); PubMed Central PMCID: PMC2927176.
58. Seike M, Goto A, Okano T, Bowman ED, Schetter AJ, Horikawa I, et al. MiR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(29):12085–90. doi: [10.1073/pnas.0905234106](https://doi.org/10.1073/pnas.0905234106) PMID: [19597153](https://pubmed.ncbi.nlm.nih.gov/19597153/); PubMed Central PMCID: PMC2715493.
59. Zhu S, Si ML, Wu H, Mo YY. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *The Journal of biological chemistry*. 2007; 282(19):14328–36. doi: [10.1074/jbc.M611393200](https://doi.org/10.1074/jbc.M611393200) PMID: [17363372](https://pubmed.ncbi.nlm.nih.gov/17363372/).
60. Miller TE, Ghoshal K, Ramaswamy B, Roy S, Datta J, Shapiro CL, et al. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *The Journal of biological chemistry*. 2008; 283(44):29897–903. doi: [10.1074/jbc.M804612200](https://doi.org/10.1074/jbc.M804612200) PMID: [18708351](https://pubmed.ncbi.nlm.nih.gov/18708351/); PubMed Central PMCID: PMC2573063.
61. Zhao JJ, Lin J, Yang H, Kong W, He L, Ma X, et al. MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *The Journal of biological chemistry*. 2008; 283(45):31079–86. doi: [10.1074/jbc.M806041200](https://doi.org/10.1074/jbc.M806041200) PMID: [18790736](https://pubmed.ncbi.nlm.nih.gov/18790736/); PubMed Central PMCID: PMC2576549.
62. Ueda R, Kohanbash G, Sasaki K, Fujita M, Zhu X, Kasthuber ER, et al. Dicer-regulated microRNAs 222 and 339 promote resistance of cancer cells to cytotoxic T-lymphocytes by down-regulation of ICAM-1. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(26):10746–51. doi: [10.1073/pnas.0811817106](https://doi.org/10.1073/pnas.0811817106) PMID: [19520829](https://pubmed.ncbi.nlm.nih.gov/19520829/); PubMed Central PMCID: PMC2705554.
63. Tan X, Zhang P, Zhou L, Yin B, Pan H, Peng X. Clock-controlled mir-142-3p can target its activator, Bmal1. *BMC molecular biology*. 2012; 13:27. doi: [10.1186/1471-2199-13-27](https://doi.org/10.1186/1471-2199-13-27) PMID: [22958478](https://pubmed.ncbi.nlm.nih.gov/22958478/); PubMed Central PMCID: PMC3482555.
64. Cheng H-YM, Papp JW, Varlamova O, Dziema H, Russell B, Curfman JP, et al. microRNA modulation of circadian-clock period and entrainment. *Neuron*. 2007; 54:813–29. doi: [10.1016/j.neuron.2007.05.017](https://doi.org/10.1016/j.neuron.2007.05.017) PMID: [17553428](https://pubmed.ncbi.nlm.nih.gov/17553428/).
65. Gattfield D, Le Martelot G, Vejnar CE, Gerlach D, Schaad O, Fleury-Olela F, et al. Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes & development*. 2009; 23:1313–26. doi: [10.1101/gad.1781009](https://doi.org/10.1101/gad.1781009) PMID: [19487572](https://pubmed.ncbi.nlm.nih.gov/19487572/).
66. Lee K-H, Kim SH, Lee HR, Kim W, Kim DY, Shin JC, et al. MicroRNA-185 oscillation controls circadian amplitude of mouse Cryptochrome 1 via translational regulation. *Molecular biology of the cell*. 2013; 24:2248–55. doi: [10.1091/mbc.E12-12-0849](https://doi.org/10.1091/mbc.E12-12-0849) PMID: [23699394](https://pubmed.ncbi.nlm.nih.gov/23699394/).
67. Tan X, Zhang P, Zhou L, Yin B, Pan H, Peng X. Clock-controlled mir-142-3p can target its activator, Bmal1. *BMC molecular biology*. 2012; 13:27. doi: [10.1186/1471-2199-13-27](https://doi.org/10.1186/1471-2199-13-27) PMID: [22958478](https://pubmed.ncbi.nlm.nih.gov/22958478/).
68. Chen R, D'Alessandro M, Lee C. miRNAs are required for generating a time delay critical for the circadian oscillator. *Current biology: CB*. 2013; 23:1959–68. doi: [10.1016/j.cub.2013.08.005](https://doi.org/10.1016/j.cub.2013.08.005) PMID: [24094851](https://pubmed.ncbi.nlm.nih.gov/24094851/).

69. Kinoshita C, Aoyama K, Matsumura N, Kikuchi-Utsumi K, Watabe M, Nakaki T. Rhythmic oscillations of the microRNA miR-96-5p play a neuroprotective role by indirectly regulating glutathione levels. *Nature communications*. 2014; 5:3823. doi: [10.1038/ncomms4823](https://doi.org/10.1038/ncomms4823) PMID: [24804999](https://pubmed.ncbi.nlm.nih.gov/24804999/).
70. Shende VR, Kim SM, Neuendorff N, Earnest DJ. MicroRNAs function as cis- and trans-acting modulators of peripheral circadian clocks. *FEBS letters*. 2014. doi: [10.1016/j.febslet.2014.05.058](https://doi.org/10.1016/j.febslet.2014.05.058) PMID: [24928439](https://pubmed.ncbi.nlm.nih.gov/24928439/).
71. Shende VR, Neuendorff N, Earnest DJ. Role of miR-142-3p in the post-transcriptional regulation of the clock gene *Bmal1* in the mouse SCN. *PloS one*. 2013; 8:e65300. doi: [10.1371/journal.pone.0065300](https://doi.org/10.1371/journal.pone.0065300) PMID: [23755214](https://pubmed.ncbi.nlm.nih.gov/23755214/).
72. Hakenberg J, Gerner M, Haeussler M, Solt I, Plake C, Schroeder M, et al. The GNAT library for local and remote gene mention normalization. *Bioinformatics*. 2011; 27(19):2769–71. Epub 2011/08/05. btr455 [pii] doi: [10.1093/bioinformatics/btr455](https://doi.org/10.1093/bioinformatics/btr455) PMID: [21813477](https://pubmed.ncbi.nlm.nih.gov/21813477/); PubMed Central PMCID: [PMC3179658](https://pubmed.ncbi.nlm.nih.gov/PMC3179658/).
73. Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, Ruch P, et al. Overview of BioCreative II gene normalization. *Genome biology*. 2008; 9 Suppl 2:S3. doi: [10.1186/gb-2008-9-s2-s3](https://doi.org/10.1186/gb-2008-9-s2-s3) PMID: [18834494](https://pubmed.ncbi.nlm.nih.gov/18834494/); PubMed Central PMCID: [PMC2559987](https://pubmed.ncbi.nlm.nih.gov/PMC2559987/).
74. Arighi CN, Lu Z, Krallinger M, Cohen KB, Wilbur WJ, Valencia A, et al. Overview of the BioCreative III Workshop. *BMC bioinformatics*. 2011; 12 Suppl 8:S1. doi: [10.1186/1471-2105-12-S8-S1](https://doi.org/10.1186/1471-2105-12-S8-S1) PMID: [22151647](https://pubmed.ncbi.nlm.nih.gov/22151647/); PubMed Central PMCID: [PMC3269932](https://pubmed.ncbi.nlm.nih.gov/PMC3269932/).
75. Giuliano KA, Johnston PA, Gough A, Taylor DL. Systems cell biology based on high-content screening. *Methods in enzymology*. 2006; 414:601–19. doi: [10.1016/S0076-6879\(06\)14031-8](https://doi.org/10.1016/S0076-6879(06)14031-8) PMID: [17110213](https://pubmed.ncbi.nlm.nih.gov/17110213/).
76. Pysalo S, Airola A, Heimonen J, Bjorne J, Ginter F, Salakoski T. Comparative analysis of five protein-protein interaction corpora. *BMC bioinformatics*. 2008; 9 Suppl 3:S6. doi: [10.1186/1471-2105-9-S3-S6](https://doi.org/10.1186/1471-2105-9-S3-S6) PMID: [18426551](https://pubmed.ncbi.nlm.nih.gov/18426551/); PubMed Central PMCID: [PMC2349296](https://pubmed.ncbi.nlm.nih.gov/PMC2349296/).
77. Tikk D, Thomas P, Palaga P, Hakenberg J, Leser U. A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature. *PLoS Comput Biol*. 2010; 6:e1000837. Epub 2010/07/10. doi: [10.1371/journal.pcbi.1000837](https://doi.org/10.1371/journal.pcbi.1000837) PMID: [20617200](https://pubmed.ncbi.nlm.nih.gov/20617200/); PubMed Central PMCID: [PMC2895635](https://pubmed.ncbi.nlm.nih.gov/PMC2895635/).
78. Segura-Bedmar I, Martinez P, de Pablo-Sanchez C. Using a shallow linguistic kernel for drug-drug interaction extraction. *Journal of biomedical informatics*. 2011; 44(5):789–804. doi: [10.1016/j.jbi.2011.04.005](https://doi.org/10.1016/j.jbi.2011.04.005) PMID: [21545845](https://pubmed.ncbi.nlm.nih.gov/21545845/).
79. Segura-Bedmar I, Martinez P, de Pablo-Sanchez C. A linguistic rule-based approach to extract drug-drug interactions from pharmacological documents. *BMC bioinformatics*. 2011; 12 Suppl 2:S1. doi: [10.1186/1471-2105-12-S2-S1](https://doi.org/10.1186/1471-2105-12-S2-S1) PMID: [21489220](https://pubmed.ncbi.nlm.nih.gov/21489220/); PubMed Central PMCID: [PMC3073181](https://pubmed.ncbi.nlm.nih.gov/PMC3073181/).
80. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*. 2012; 40:D13–25. doi: [10.1093/nar/gkr1184](https://doi.org/10.1093/nar/gkr1184) PMID: [22140104](https://pubmed.ncbi.nlm.nih.gov/22140104/).
81. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006; 22:1600–7. PMID: [16606683](https://pubmed.ncbi.nlm.nih.gov/16606683/).
82. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nature genetics*. 2004; 36:431–2. doi: [10.1038/ng0504-431](https://doi.org/10.1038/ng0504-431) PMID: [15118671](https://pubmed.ncbi.nlm.nih.gov/15118671/).
83. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic acids research*. 2009; 37:D793–6. doi: [10.1093/nar/gkn665](https://doi.org/10.1093/nar/gkn665) PMID: [18842627](https://pubmed.ncbi.nlm.nih.gov/18842627/).
84. Gattiker A, Michoud K, Rivoire C, Auchincloss AH, Coudert E, Lima T, et al. Automated annotation of microbial proteomes in SWISS-PROT. *Computational biology and chemistry*. 2003; 27:49–58. PMID: [12798039](https://pubmed.ncbi.nlm.nih.gov/12798039/).
85. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000; 25:25–9. doi: [10.1038/75556](https://doi.org/10.1038/75556) PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/).
86. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000; 28:27–30. PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/).