



UNIVERSIDADE FEDERAL DE ALAGOAS – UFAL
INSTITUTO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE – ICBS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE- PPGCS

JOSÉ LUIZ ARAUJO SANTOS

**INVESTIGAÇÃO DE NOVOS GENES CANDIDATOS A REGULADORES DO
RITMO CIRCADIANO IDENTIFICADOS A PARTIR DE MINERAÇÃO DE DADOS
DE EXPRESSÃO GÊNICA EM CERÉBRO DE MAMÍFEROS**

MACEIÓ

2017

JOSÉ LUIZ ARAUJO SANTOS

**INVESTIGAÇÃO DE NOVOS GENES CANDIDATOS A REGULADORES DO
RITMO CIRCADIANO IDENTIFICADOS A PARTIR DE MINERAÇÃO DE DADOS
DE EXPRESSÃO GÊNICA EM CERÉBRO DE MAMÍFEROS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Saúde da Universidade Federal de Alagoas como requisito parcial para obtenção do título de Mestre em Ciências da Saúde.

Orientador (a): Profº Dr. Tiago Gomes de Andrade

MACEIÓ

2017

Catlogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário Responsável: Valter dos Santos Andrade

S237a

Santos, José Luiz Araujo.

Investigação de novos genes candidatos a reguladores do ritmo circadiano identificados a partir de mineração de dados de expressão gênica em cérebro de mamíferos / José Luiz Araujo Santos. – 2017. 100 f. ; il.

Orientador: Tiago Gomes de Andrade.

Dissertação (Mestrado em Ciências da Saúde) – Universidade Federal de Alagoas. Instituto de Ciências Biológicas e da Saúde. Programa de Pós-Graduação em Ciências da Saúde. Maceió, 2017.

Bibliografia: f. 95-100.

1. Ritmos circadianos. 2. Mineração de dados. 3. Cérebro. 4. Gene. 5. Cerebelo. 6. Expressão gênica. I. Título.

CDU: 577.21/.218



Universidade Federal de Alagoas
Instituto de Ciências Biológicas e da Saúde
Programa de Pós-graduação em Ciências da Saúde

ICBS - UFAL – Campus A. C. Simões
Av. Lourival Melo Mota, S/N
Cidade Universitária – Maceió-AL
CEP: 57072-900
E-mail: ppgcs9@gmail.com
Fone: 82 3214 1850


Folha de Aprovação

Jose Luiz Araujo Santos


Investigação de novos genes candidatos a reguladores do ritmo circadiano identificados a partir de mineração de dados de expressão gênica em cérebros de mamíferos

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Ciências da Saúde da Universidade Federal de Alagoas e aprovada em 06 de março de 2017.

Banca Examinadora



Prof. Dr. Tiago Gomes de Andrade (Orientador)



Prof. Dr. Daniel Leite Goes Gitaí – (UFAL)



Prof. Dr. Pollyana Almeida Costa dos Santos - (UNCISAL)

*“Dedico aos meus pais,
Valdenice e José Luiz, e in
memoriam a minha avó,
Aurora.*

AGRADECIMENTOS

“Não conquistaria o que conquistei até hoje, se as pessoas aqui listadas e lembradas não tivessem contribuído e acrescentado em minha vida!” Desta forma, dedico-lhes meus sinceros agradecimentos;

Primeiramente, agradeço aos meus pais, Valdenice Brito de Araujo e José Luiz Martins, às minhas irmãs, Maria da Conceição e Deisiane, pelo amor incondicional, pelos conselhos, pelo incentivo e força, e por compartilhar também dos meus sonhos.

Agradeço imensamente, a minha avó, Aurora (*in memória*), que já não faz mais parte deste mundo, mas sempre torceu pelo meu crescimento.

Agradeço a minha segunda família, amigos e companheiros de pesquisa do Grupo LABMEG (Laboratório de Biologia Molecular e Expressão Gênica). Ao Daniel Coimbra e Diego Siqueira, pelos momentos divertidos com direito a fotos com diferentes caras e bocas (arquivos não publicados), pelos conselhos dados, pelas lamentações compartilhadas, pelos apelidos com fins de entretenimento, pelas caronas dadas da UFAL até minha casa, e pelas contribuições e ajuda nesta pesquisa. À Mayara Barbosa e Ellyda Lopes, Denise Macêdo e Ana Carolina pela amizade dentro e fora dos muros da UFAL, pelo companheirismo nos momentos bons e ruins. À Maria Oliveira, sempre pelos conselhos, ressalto que a tratamos como “mãe Maria”. E aos demais integrantes deste grupo.

Agradeço bastante a Luana Karen, pela ajuda na coleta e armazenamento dos dados adquiridos do portal *Allen Brain Atlas*.

Agradeço imensamente ao meu orientador, o **Profº Drº** Tiago Gomes de Andrade, que fundou e consolidou o do Grupo LABMEG, que me deu a oportunidade de conhecer o mundo científico ainda na graduação com o primeiro PIBIC, que oportunizou a minha entrada no mestrado, que me fez ver que a ciência deve ser levada a sério, com dedicação, responsabilidades, compromisso e muito estudo. Ele também já me deu bastantes conselhos e broncas, que me fizeram crescer como pessoa e como pesquisador. Além disso, sempre se dispôs discutir dados, novas ideias, resultados, a corrigir os meus manuscritos incontáveis vezes, em períodos de fim de semana e feriados, inclusive, nesse carnaval. Por receber minhas mensagens e áudios de voz no *Whatsapp* nas madrugadas e não reclamar. Meu muito obrigado pela paciência, ensinamentos e disponibilidade;

Agradeço às minhas verdadeiras amigas, Eliane Paz, Iolanda Lima, Luana Ribeiro, Eliane Midian, Gerlane Barros, pois sei que torcem por mim todos os dias;

Agradeço aos meus amigos, Felipe Jackson, Fernando Mizael, Ademir Ferreira, Dhoone Menezes, Thiago Dantas, Alex Ranière, Junior Belém e Jonathan Garcia, por ouvirem minhas lamentações, pelos conselhos de amigos, pela cumplicidade e os encontros casuais de diversões;

Agradeço a Demytre Messias, Dona Iracilda e seu esposo, por me receber gentilmente e me hospedar em sua casa, todas as vezes que precisei ir a Maceió para resolver questões do mestrado.

Agradeço a todos os meus professores, desde a Educação Básica a Pós-graduação, cada um teve sua contribuição neste processo.

Agradeço à Universidade Federal de Alagoas – UFAL, ao Instituto de Ciências Biológicas e da Saúde – ICBS, ao Programa de Pós-Graduação em Ciências da Saúde- PPGCS e a todos os servidores que fazem com que estas unidades funcionem, especialmente à Professora Salete Smannioto, coordenadora do programa, por atender as nossas necessidades enquanto pós-graduandos, ao técnico Jonathan, por resolver as questões burocráticas e ao técnico Rafael, do laboratório de ensino em genética, por facilitar o acesso ao laboratório, aos equipamentos e utilizados nas minhas aulas práticas de Estágio e Docência.

Agradeço à Capes, Fapeal e UFAL pelos financiamentos de bolsas e projetos.

Agradeço à banca de qualificação, composta pelos professores Marcelo Duzzione e Adriana Ximines, pela contribuição científica e na escrita deste manuscrito.

Agradeço à banca de defesa, composta pelos professores Pollyana Almeida e Daniel Gitaí, por aceitarem o convite, pelas correções, contribuições e sugestões neste trabalho.

Por fim, agradeço à vida, pela oportunidade de viver.

EPÍGRAFE

“É como se minha vida fosse uma pintura inacabada. E, como artista dessa pintura, eu preciso preencher todas as lacunas, e torná-la bonita outra vez”.

Stefani J. A.

RESUMO

Os ritmos circadianos são fenômenos biológicos que se repetem com frequência de aproximadamente 24 horas. Em nível molecular, esses fenômenos cíclicos são controlados por um conjunto de genes centrais, denominados de genes circadianos. O envolvimento dos genes circadianos em mamíferos começou a ser compreendido a partir dos anos 90, com a descoberta primeiramente do gene, *Clock*, posteriormente de *Bmal1*, *Cry1*, *Cry2*, *Per1*, *Per2* e *Per3*, *Rora*, *Rorb*, *Rorc*, *Nr1d1* e *Nr1d2*, dentre outros. Diferentes abordagens e métodos têm sido utilizados na identificação de genes envolvidos com o ritmo circadiano. Uma abordagem mais atual, conhecida como mineração de dados, surgida a partir da bioinformática, definida como uma análise de grandes conjuntos de dados com o objetivo de encontrar novas relações entre eles, resumindo-os em forma compreensíveis e úteis para gerar novos conhecimentos, tem sido utilizada no descobrimento dessas moléculas funcionais e do seu envolvimento esse sistema de temporização. O gene *Chrono*, por exemplo, foi identificado como um importante componente do sistema de temporização circadiano em mamíferos, através de uma abordagem computacional, baseada na mineração de dados. Nesta pesquisa, utilizamos dados de expressão gênica em larga escala do cerebelo humano e murino, disponíveis no portal *Allen Brain Atlas* como um método preditivo para a identificação de genes que possuem oscilação ou função circadiana, por meio da correlação (r de *Pearson*) de expressão espacial com 19 principais genes circadianos, (*Bmal1*, *Bmal2*, *Clock*, *Cry1*, *Cry2*, *Csnk1a*, *Csnk1d*, *Csnk1e*, *Fblx3*, *Fblx21*, *Npas2*, *Nr1d1*, *Nr1d2*, *Per1*, *Per2*, *Per3*, *Rora*, *Rorb*, e *Rorc*). 2% dos genes com os valores extremos de correlações nas duas espécies foram confrontados com dados em larga escala de estudos com transcriptoma, proteoma, ChiP-seq, interação proteína-proteína e RNAi e analisadas suas interações funcionais através do portal *on line STRING*. Nossos resultados demonstram que 19.907 transcritos no cerebelo de *M. musculus* e 29.176 transcritos no cerebelo de *H. sapiens*, possuem perfis de expressão espacial correlacionados com um ou mais dos 19 genes circadianos investigados, sendo 14.242 genes comuns entre as duas espécies. A distribuição dos dados de correlação considerando os genes que apresentaram ou não oscilação circadiana previamente identificada indica um perfil onde o conjunto de genes que apresentaram variação também tendem a apresentar maiores valores de correlação, positiva ou negativa, para ambas as espécies. A comparação das médias dos valores $de r$ em módulo ($|r|$) destes grupos indicou valores significativamente maiores para o grupo que apresenta variação circadiana ($p < 0,0001$) para os dados de camundongos. Além disto, os 2% dos genes com maiores valores de r positivos e negativos apresentou uma frequência maior de genes com variação circadiana em quase todos os estudos circadianos analisados, exceto no estudo de RNAi, para *M. musculus* e *H. sapiens*, comparativamente ao grupo de 2% dos genes com valores de r mais próximos de zeros negativos e positivos. Os 100 genes correlacionados com os maiores valores de $|r|$ obtidos para *M. musculus* foram ranqueados com base na soma destes valores com os valores de correlação obtidos para o gene homólogo correspondente em *H. sapiens*, a fim de selecionar genes candidatos para futuros ensaios experimentais. O gene *Cartpt* ocupa a primeira posição do quadro, apresentando o maior resultado da somatória entre os valores de $|r|$ obtidos para *M. musculus* e *H. sapiens*, sendo forte candidato a controlar o ritmo alimentar em cerebelo de mamíferos. Este gene foi correlacionado com o ritmo circadiano alimentar em camundongos. *Per2*, *Cry2*, *Bmal1* e *Rora*, quatro genes circadianos estão presentes nesta lista. *Gpr149*, *Ptprp*, *Nipal4* estão entre os 10 mais bem ranqueados e não apresentam nenhuma evidência prévia de envolvimento com o ritmo, constituindo-se como novos candidatos a genes circadianos em

cerebelo de mamíferos. A estratégia utilizada tem potencial para identificar genes importantes para o ritmo circadiano em estruturas específicas do cérebro humano e de camundongos. No entanto, análises de expressão gênica e ensaios funcionais são necessários para validar o envolvimento destes genes candidatos.

Palavras-Chave: Ritmos Circadianos. Mineração de Dados. Cérebro. Gene. Cerebelo. Expressão Gênica.

ABSTRACT

Circadian rhythms are a biological phenomena that is repeated with frequency of approximately 24 hours. At the molecular level, these cyclic phenomena are controlled by a set of central genes called clock genes. The involvement of clock genes and clock-controlled genes in the circadian rhythms of mammals began to be understood since the 90s, with the discovery of the genes: *Clock*, *Bmal1*, *Cry1*, *Cry2*, *Per1*, *Per2* and *Per3*, *Rora*, *Rorb*, *Rorc*, *Nr1d1* and *Nr1d2*. Different approaches and methods have been used in the identification of genes involved with the circadian rhythm. A more current approach known as data mining arose from bioinformatics, and it is defined as an analysis of large data sets with the objective of finding new relationships between them, summarizing them comprehensibly and usefully in generating new knowledge, it has been used in the discovery of these functional molecules and their involvement in this timing system. The *Chrono* gene, for example, has been identified as an important component of the circadian timing system in mammals, using a computational approach based on data mining. In this study, we used gene expression data in wide range of human and murine cerebellum, available via the *Allen Brain Atlas* portal as a predictive method for the identification of genes that have circadian oscillation or function, through the correlation of spatial expression (*Pearson's r*) with 19 major circadian genes (*Bmal1*, *Bmal2*, *Clock*, *Cry1*, *Cry2*, *Csnk1a*, *Csnk1d*, *Csnk1e*, *Fblx3*, *Fblx21*, *Npas2*, *Nr1d1*, *Nr1d2*, *Per1*, *Per2*, *Per3*, *Rora*, *Rorb* and *Rorc*). 2% of the genes with the extreme values of correlations in the two species were compared with data on a large scale studies with transcriptoma, proteoma, ChiP-seq, RNAi and protein-protein interaction, and their functional interactions were analyzed through the online portal STRING. Our results show that 19,907 genes in the cerebellum of *Mus musculus* and 29,176 genes in the cerebellum of *Homo sapiens* have spatial expression profiles related to one or more of the 19 circadian genes investigated and 14,242 genes common to both species. The distribution of correlation data considering the genes that present or not circadian variation previously identified shows a profile where the number of genes that exhibit changes also tend to have larger correlation values, positive or negative, to both species. The comparison of average values of *r* in module ($|r|$) of these groups showed significantly higher values for the group that shows circadian variation ($p < 0.0001$) for the mice data. Furthermore, 2% of the genes with larger values of $|r|$ showed a higher frequency of genes with circadian variation in almost all circadian studies analyzed, except in the study of RNAi, for *M. musculus* and *H. sapiens*, compared to the group of 2% of the genes with *r* values closer to negative and positive zeros. The 100 genes related to higher values of $|r|$ obtained for *M. musculus* were ranked based on the sum of these values with the correlation values obtained for the corresponding homologous gene in *H. sapiens* in order to select candidate genes for future experimental studies. The *Cartpt* gene occupies the first position of the chart, with the highest result of the sum of the values of $|r|$ obtained for *M. musculus* and *H. sapiens*, being a strong candidate to control the feeding rhythm in mammalian cerebellum. This gene has been associated with food circadian rhythm in mice. *Per2*, *Cry2* and *Rora*, three circadian genes are on this list. *Gpr149*, *Ptprp*, *Nipal4* are among the 10 best-ranked and present no prior evidence of involvement with the rhythm, constituting as new candidates for circadian genes in mammalian cerebellum. The strategy used has the potential to identify important genes for the circadian rhythm in specific structures of the human and mice's brain. However, analysis of gene expression and functional assays are needed to validate the involvement of these candidate genes.

Keywords: Circadian Rhythms. Data mining. Brain. Gene. Cerebellum. Gene Expression.

LISTA DE FIGURAS

Figura 01. Esquema simplificado do sistema de temporização circadiano de mamíferos.....	20
Figura 02. Alça de retroalimentação molecular transcricional e traducional do sistema circadiano em mamíferos.....	22
Figura 03. O envolvimento dos mecanismos pós-transcricionais na regulação dos ritmos circadianos.....	23
Figura 04. Potenciais osciladores periféricos no cérebro de mamíferos e o oscilador central.....	26
Figura 05. Página principal do Portal <i>Allen Brain Atlas</i>	32
Figura 06. Expressão do gene relógio <i>Per1</i> em diversas regiões do cérebro de camundongos.....	35
Figura 07. Mapa quente demonstrando a expressão espacial do gene <i>Per1</i> no cérebro de humanos.....	36
Figura 08. Fluxograma simplificado das etapas para identificação de genes candidatos ao controle dos ritmos circadianos em mamíferos.....	42
Figura 09. Fluxograma ilustrando as etapas seguidas na busca e armazenamento dos dados de genes que apresentam expressão espacial correlacionada com 16 genes relógio, em cerebelo das espécies <i>M. musculus</i> e <i>H. sapiens</i> , no portal <i>Allen Brain Atlas</i> , nas ferramentas <i>mouse brain</i> e <i>human brain</i>	44
Figura 10. Distribuição dos 19.907 genes correlacionados com os genes circadianos no cerebelo de <i>M. musculus</i> , de acordo com seus valores de <i>r</i> que variam de -0,633 a 0,729.....	51
Figura 11. Distribuição dos 29.176 genes correlacionados com os genes circadianos no cerebelo de <i>H. sapiens</i> , de acordo com seus valores de <i>r</i> que variam de - 0,899 a 0,948.....	52
Figura 12. Distribuição por Frequência relativa (em relação ao total de cada grupo) dos valores de correlação de <i>Pearson</i> de 19 genes circadianos com expressão correlacionada entre si e do total de genes correlacionados com 19 genes circadianos obtidos em cerebelo de <i>M. musculus</i> , por faixas de 0,1.....	55
Figura 13. Distribuição por Frequência relativa (em relação ao total de cada grupo) dos valores de correlação de <i>Pearson</i> de 19 genes circadianos com expressão correlacionada entre si e do total de genes correlacionados com 19 genes circadianos obtidos em cerebelo de <i>H. sapiens</i> , por faixas de 0,1.....	56
Figura 14. Frequência relativa dos genes correlacionados obtidos para espécie <i>M. musculus</i> por faixa dos valores de <i>r</i> , que apresentam ou não oscilação circadiana de acordo com o trabalho de ZANGH et al.,2014.....	61

- Figura 15.** Distribuição em porcentagem dos genes correlacionados obtidos para espécie *H. Sapiens* por faixa dos valores de r , que apresentam ou não oscilação circadiana de acordo com o trabalho de ZANGH et al.,2014.....61
- Figura 16.** Comparação das médias em módulo dos valores de $|r|$, correspondentes aos genes que, conforme os dados de ZHANG et al.,2014, que apresentaram ou não oscilação circadiana em cerebelo de dados obtidos para a espécie *M. musculus* (teste de Mann Whitney $p < 0,0001$).....62
- Figura 17.** Comparação das médias em módulo dos valores de $|r|$, correspondentes aos genes que, conforme os dados de LI et al.,2013, que apresentaram ou não oscilação circadiana em cerebelo de dados obtidos para a espécie *H. Sapiens* (teste de Mann Whitney $p < 0,0001$).....62
- Figura 18.** Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebelo da espécie *M. musculus*, que apresentam os maiores e os menores valores de r positivos, nos estudos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi.....63
- Figura 19.** Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebelo da espécie *M. musculus*, que apresentam os maiores e os menores valores de r negativos, nos estudos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi.....64
- Figura 20.** Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebelo da espécie *H. Sapiens*, que apresentam os maiores e os menores valores de r positivos, nos estudos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi.....65
- Figura 21.** Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebelo da espécie *H. Sapiens*, que apresentam os maiores e os menores valores de r negativos, nos estudos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi.....65
- Figura 22.** Análise de interações funcionais das 19 proteínas circadianas estudadas com 397 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *M. musculus*, por meio do portal STRING (<http://string-db.org/>).....67
- Figura 23.** Análise de interações funcionais das 19 proteínas circadianas estudadas com 397 proteínas, correspondentes aos 2% dos genes com os menores valores de correlações positivas

e negativas em cerebelo de *M. musculus*, por meio do portal STRING (<http://string-db.org/>).....69

Figura 24. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de interações funcionais das 416 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *M. musculus*, mais as 19 proteínas circadianas e 604 interações, comparados as 416 proteínas com os menores valores de correlações negativos e positivos, somadas às 19 proteínas circadianas, e as 312 interações.....70

Figura 25. Análise de interações funcionais das 19 proteínas circadianas estudadas com 621 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *H. sapiens*, por meio do portal STRING (<http://string-db.org/>). A área em destaque na figura refere-se à localização das 19 proteínas circadianas estudadas.....71

Figura 26. Análise de interações funcionais das 19 proteínas circadianas estudadas com 621 proteínas, correspondentes aos 2% dos genes com os menores valores de correlações positivas e negativas em cerebelo de *H. sapiens*, por meio do portal STRING (<http://string-db.org/>).....72

Figura 27. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de interações funcionais das 621 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *H. sapiens* mais as 19 proteínas circadianas e 804 interações, comparados as 621 proteínas com os menores valores de correlações negativos e positivos, somadas às 19 proteínas circadianas, e as 138 interações.....72

Figura 28. Imagens obtidas através do portal *Allen Brain Atlas*, do padrão de expressão espacial dos genes *Rorb*, *Cartp* e *Ppp3r1* através da técnica de *HIS* no cérebro de *M. musculus*.....79

Figura 29. Imagens obtidas através do portal *Allen Brain Atlas*, do padrão de expressão espacial dos genes *Cry2*, *Cartp*, *Cry1* e *Ppp3r1* através da técnica de *Mircroarray* no cérebro de *H. Sapiens*.....80

Figura 30. Percentual de evidências de 100 genes candidatos identificados com base na correlação de expressão espacial no cérebro de *M. musculus* e *H. sapiens*, a partir de dados disponíveis no *Allen Brain Atlas*, nos estudos circadianos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi.....81

LISTA DE TABELAS

Tabela 01. Os principais recursos públicos do portal <i>Allen Brain Atlas</i> , com suas respectivas características e ordenados de forma cronológica de desenvolvimento.....	33
Tabela 02. Valores de correlação de <i>Pearson</i> (r) de 10 genes com expressão correlacionada com os 19 genes circadianos, em cerebelo de camundongos	53
Tabela 03. Valores de correlação de <i>Pearson</i> (r) de 10 genes com expressão correlacionada com os 19 genes circadianos, em cerebelo de humanos	54
Tabela 04. Valores de correlação de <i>Pearson</i> (r) de 19 genes circadianos com expressão correlacionada entre si, em cerebelo de camundongos	58
Tabela 05. Valores de correlação de <i>Pearson</i> (r) de 10 genes com expressão correlacionada com os 19 genes circadianos, em cerebelo de humanos	59
Tabela 06. 100 genes correlacionados com os maiores valores de $ r $ obtidos para <i>M. musculus</i> ranqueados com base na soma dos valores de correlação obtidos para o gene homólogo em <i>H. sapiens</i> , além de dados experimentais de estudos cronobiológicos	74

LISTA DE ABREVIATURAS

ABA - Allen Brain Atlas
AMPK - Amp Kinase
Arntl - Aryl Hydrocarbon Receptor Nuclear Translocator Like
AVP - Arginina Vasopressina
Bhlhe40 - Basic Helix-Loop-Helix Family, Member E 40
Bhlhe41 - Basic Helix-Loop-Helix Family, Member E 41
Bmal1 - Brain And Muscle-Arnt-Like Protein-1
CAMK II - Palmodulin-Dependent Kinase
CCGs - Clock Controlled Genes
Chip-Seq - Chromatin Immunoprecipitation Sequencing
Clock - Circadian Locomotor Output Cycles Kaput
CK2 - Casein Kinase 2
Ck1 ϵ - Casein Kinase 1, Épsilon
Ck1 δ - Casein Kinase 1, Delta
Cry 1 - Cryptochrome1
Cry2 - Cryptochrome 24
Fbx13 - F-Box And Leucine Rich Repeat Protein 3
Fbx121 - F-Box And Leucine Rich Repeat Protein 21
GRP - Peptídeo Liberador De Gastrina
HIS- Hibridização In Situ
Id – Identificação
IMS-PCR - Intra-Module Scanning-Polymerase Chain Reaction
MAPK - Mitogen-Activated Protein Kinase
Npas2 - Neuronal Pas Domain Protein 2
NSQ - Núcleo Supraquiasmático
Nr1d1 - Nuclear Receptor Subfamily 1, Group D, Member 1
Nr1d2 - Nuclear Receptor Subfamily 1, Group D, Member 2
Per 1 - Period1
Per 2 - Period2
Per 3 - Period3
PKA - Protein Kinase A
Pubmed - Biblioteca Nacional De Medicina Dos Estados Unidos
RNAi - Rna de Interferência
RNA m - Rna Mensageiro
Rora - Rar-Related Orphan Receptor Alpha
Rorb - Rar-Related Orphan Receptor Beta
Rorc - Rar-Related Orphanreceptor Gamma
RT-PCR - Real Time Polymerase Chain Reaction
SCF - Skp1-Cullin-F-Box Protein
TRH - Trato Retino Hipotalâmico
VIP - Polipeptídeo Intestinal Vasoativo

Sumário

1.	INTRODUÇÃO	18
2.	REVISÃO DA LITERATURA.....	19
2.1.	RITMOS CIRCADIANOS	19
2.2.	MECANISMOS MOLECULARES DO RELÓGIO CIRCADIANO EM MAMÍFEROS	22
2.3	RITMOS CIRCADIANOS NO CÉREBRO	25
2.3.1.	O RELÓGIO CENTRAL	25
2.3.2.	RELÓGIOS PERIFÉRICOS.....	26
2.5	MINERAÇÃO DE DADOS NA BUSCA DE NOVOS GENES CANDIDATOS AO CONTROLE DOS RITMOS CIRCADIANOS EM MAMIFEROS.....	31
2.6	ALLEN BRAIN ATLAS.....	32
2.6.1	ALLEN MOUSE BRAIN ATLAS	36
2.6.2	ALLEN HUMAN BRAIN ATLAS.....	37
2.6.3	EXPRESSÃO ESPACIAL DE GENES CORRELACIONADOS.....	38
3.0	JUSTIFICATIVA	38
4.	OBJETIVOS	40
4.1	OBJETIVO GERAL.....	40
4.2	OBJETIVOS ESPECÍFICOS.....	40
5.	MATERIAL E MÉTODOS	41
5.1	MINERAÇÃO DE DADOS.....	41
5.1.1	SELEÇÃO DE GENES CORRELACIONADOS AO RITMO CIRCADIANO NO CEREBELO HUMANO E DE CAMUNDONGO	41
5.2.1	DADOS DE TRANSCRIPTOMA	46
5.2.2	DADOS DE PROTEOMA	47
5.2.3	DADOS DE <i>CHIP-SEQ</i>	48
5.2.4	DADOS DE INTERAÇÃO PROTEÍNA-PROTEÍNA	49
5.2.5	DADOS DE RNA DE INTERFERÊNCIA (RNAI).....	49
5.3	ANÁLISE DE REDES DE INTERAÇÃO FUNCIONAL PROTEÍNA-PROTEÍNA.....	49
5.4	ANÁLISES ESTATÍSTICAS	50
6	RESULTADOS.....	51
6.1	IDENTIFICAÇÃO DE GENES CORRELACIONADOS A GENES CIRCADIANOS NO CEREBELO DE HUMANOS E DE CAMUNDONGOS A PARTIR DO <i>ALLEN BRAIN ATLAS</i>	51
6.2	ANÁLISE DA CORRELAÇÃO ENTRE OS 19 GENES CIRCADIANOS EM CEREBELO DE <i>MUS MUSCULUS</i> E <i>H. SAPIENS</i>	56
6.3	VALIDAÇÃO DOS DADOS COM BASE EM ESTUDOS DE TRANSCRIPTOMA CIRCADIANO DE CEREBELO	61

7. DISCUSSÃO	83
7.1 GENES CIRCADIANOS CANDIDATOS IDENTIFICADOS EM CEREBELO.....	91
8. CONCLUSÕES.....	93
9. REFERÊNCIAS.....	95

1. INTRODUÇÃO

Os ritmos circadianos são fenômenos biológicos presentes em quase todos os organismos e que se repete com frequência de aproximadamente 24 horas, sendo regulados por pistas ambientais, principalmente a luz. Esses ritmos são controlados sistemicamente pelos núcleos supraquiasmáticos localizados na base do hipotálamo, conhecidos também como relógio ou marca-passo central. Outras estruturas, entretanto, no cérebro e em outros órgãos, também apresentam ritmos autônomos, como o fígado, o rim, o hipocampo, corpo estriado e cerebelo, dentre outros, conhecidos como relógios periféricos (HASTINGS et al., 2003; MARQUES e BARRETO, 2003; KULLER, Rikard, 2002). Em nível molecular, esses fenômenos cíclicos são controlados por um conjunto de genes centrais, denominados de genes circadianos, altamente conservados e ubiquamente expressos nos diferentes tecidos e células. Mecanismos transcricionais, traducionais e pós-transcricionais/traducionais participam da geração de ritmos moleculares que se repercutem nos vários ritmos fisiológicos e comportamentais observados (MOHAWK, Jennifer A.; GREEN, Carla B.; TAKAHASHI, Joseph S., 2012; KOJIMA, Shihoko; SHINGLE, Danielle L.; GREEN, Carla B., 2011).

O envolvimento dos genes circadianos e genes controlados pelo relógio nos ritmos circadianos de mamíferos começou a ser compreendido a partir dos anos 90, com a descoberta primeiramente do gene *Clock* e posteriormente dos demais como: *Bmal1*, *Cry1*, *Cry2*, *Per1*, *Per2* e *Per3*, *Rora*, *Rorb*, *Rorc*, *Nr1d1* e *Nr1d2* e outros. Estes genes e suas respectivas proteínas fazem parte de uma alça molecular de retroalimentação que inibe e ativa seus próprios processos de transcrição, assim como a de milhares de genes controlados pelo relógio, e possuem suas funções experimentalmente comprovadas no envolvimento com estes ritmos (VITATERNA, M. Hotz et al., 1994; TAKUMI, Toru et al., 1998; ZYLKA, Mark J. et al., 1998; BUNGER, Maureen K. et al., 2000).

Com o advento de técnicas modernas de análises em larga escala de transcriptoma e proteoma, sabe-se atualmente que a depender do tecido e estrutura cerca de até 43% dos transcritos possuem expressão, e em nível proteico, pelo menos 20% das proteínas também possuem expressão circadiana, entretanto, o papel que grande parte desses genes e dessas proteínas exerce sobre o sistema molecular de temporização circadiana de mamíferos ainda precisa ser investigado. Além disso, limitações metodológicas tem resultado em

inconsistências nos dados produzidos sobre o conjunto de genes que estaria envolvido com ritmos circadianos nos diversos tecidos (ZHANG et al., 2014; LI et al. 2013; MAUVOISIN et al., 2014; CHIANG et al., 2014; ROBLES; COX; MANN, 2014).

Recentemente, ferramentas de bioinformática e mineração de dados têm sido bastante utilizadas no descobrimento dessas moléculas funcionais e do seu envolvimento nesse sistema de temporização (HAND, David J.; MANNILA, Heikki; SMYTH, Padhraic., 2001). O gene *Chrono*, por exemplo, foi identificado como um importante componente do sistema de temporização circadiano em mamíferos, através de uma abordagem computacional (ANAFI, Ron C. et al., 2014). Outro estudo identificou 118 potenciais genes controlados pelo relógio, que estão envolvidos com a regulação circadiana dos ritmos em mamíferos, através de uma análise feita por um *pipeline* de bioinformática que integra diferentes fontes e tipos de informações de bancos de dados LEHMANN et al., 2015). Um terceiro estudo utilizou 1000 genes encontrado nos resultados de ANAFI e colaboradores, (2014), para confrontar com dados de larga escala, de interação proteína-proteína, ChiP-seq e proteoma, encontrando 11 potenciais genes candidatos envolvimento com os ritmos circadianos (BHARGAVA et al., 2015).

Nesta pesquisa, utilizamos dados de expressão gênica em larga escala do cérebro humano e murino, disponíveis no portal *Allen Brain Atlas* como um método preditivo para a identificação de genes que possuem oscilação ou função circadiana, por meio da correlação de expressão espacial com os principais genes circadianos. Confrontamos os resultados obtidos com dados em larga escala de estudos circadianos em transcriptoma, proteoma, ChiP-seq, RNAi e interação proteína-proteína, como uma estratégia inicial de validação dos dados.

2. REVISÃO DA LITERATURA

2.1. RITMOS CIRCADIANOS

Ritmos biológicos são eventos que se repetem regularmente, estão associados a ciclos ambientais e sua determinação endógena é comprovada cientificamente (MARQUES e MENNA-BARRETO, 2003). Estes ritmos são encontrados em diferentes organismos vivos, sejam eles os mais simples, seres unicelulares, como cianobactérias, ou mais complexos, como plantas e animais (LEMMER, Björn 2009; LOWREY; TAKAHASHI, 2004; BELL-PEDERSEN et al., 2005). Os ritmos biológicos são influenciados por fatores ou

sincronizadores ambientais externos denominados *Zeitgebers* (termo de origem germânica que significa “doadores de tempo”) tais como a luz, a temperatura, a disponibilidade de alimentos entre outros. Entretanto, esses ritmos persistem na ausência destes fatores ambientais, comprovando a existência de ritmos biológicos gerados endogenamente (MARQUES e MENNA-BARRETO, 2003; PEREIRA et al., 2009).

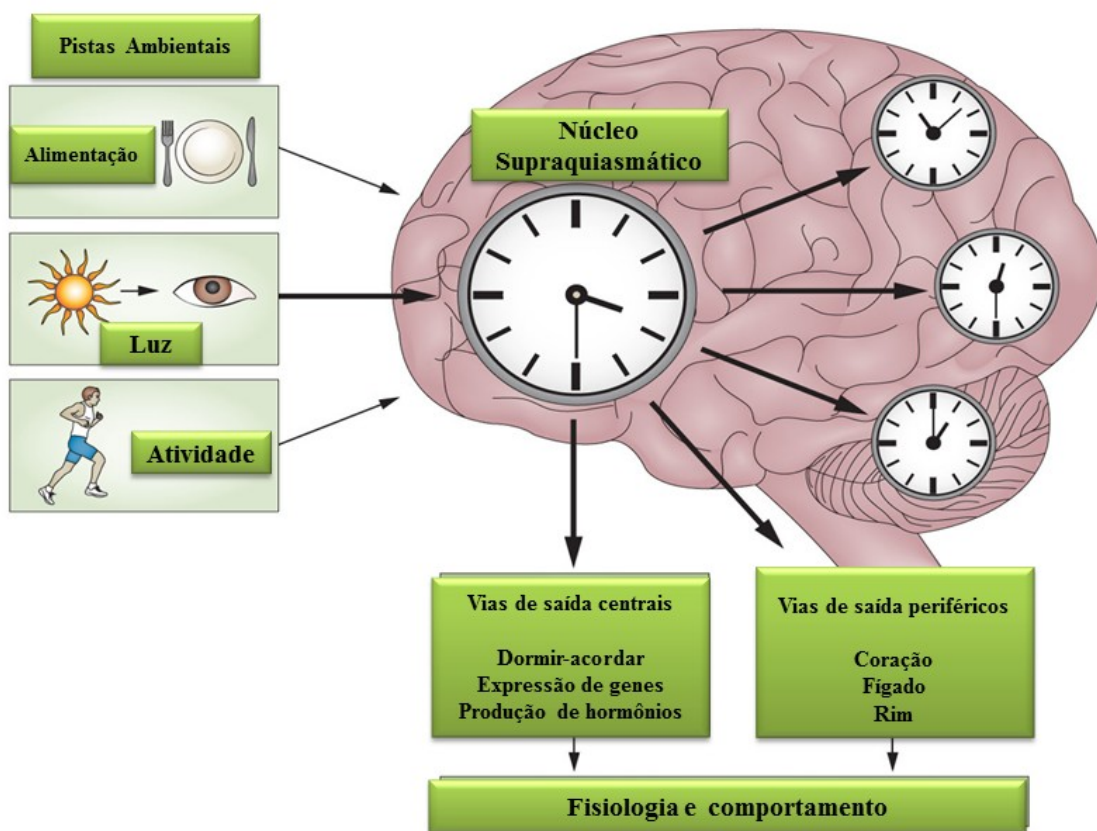
Os ritmos biológicos apresentam diferentes padrões cíclicos e compreendem três categorias de ritmos que diferem em seus períodos: ritmos circadianos ou diários (do grego *circa*, que significa aproximado, e *dian* – dia), que duram aproximadamente 24h, com variação de mais ou menos 4h. Exemplos de ritmos circadianos são: o ciclo sono/vigília em humanos, atividade e repouso em roedores, variações diárias da temperatura, comportamento alimentar, expressão gênica e produção de hormônios (HASTINGS et al., 2003; MARQUES e BARRETO, 2003; KULLER, Rikard, 2002). Os ritmos ultradianos possuem um período inferior à 20h, como o ciclo respiratório, batimentos cardíacos e pulsação sanguínea. Os ritmos infradianos, que possuem período superior à 28h, tem como exemplos o ciclo menstrual, hibernação e movimentos migratórios (BOTBOL, Michel, et al., 2003).

Estes ritmos nos vertebrados são sincronizados aos ritmos ambientais a partir de um sistema composto por uma via de entrada (*input*), onde sinais ambientais, tais como: alimentação, temperatura e principalmente a luz, são captados, sincronizando a atividade do núcleo supraquiasmático (NSQ), também conhecido como o relógio biológico ou marca-passo central, localizado na base do hipotálamo no cérebro de mamíferos. (MARQUES e BARRETO 2003; BUHR e TAKAHASHI, 2013; LOWREY e TAKAHASHI, 2004; TOSINI, Gianluca et al., 2008). Os sinais gerados no NSQ constituem a via de saída (*output*), e regulam sistemicamente e ritmicamente os diversos eventos bioquímicos, fisiológicos e comportamentais essenciais a sobrevivência dos seres vivos. (Lowrey e Takahashi, 2011) (Figura 01).

O sistema de temporização circadiano de mamíferos é composto não somente pelo NSQ, mas também por outros relógios biológicos, presentes em diversas células e tecidos no organismo (DIBNER, Charna; SCHIBLER, Ueli; ALBRECHT, Urs, 2010). De fato, evidências mostram que os ritmos moleculares estão presentes mesmo em células individuais (LIU, Andrew C.; LEWIS, Warren G.; KAY, Steve A., 2007). Estes osciladores interagem e são sincronizados pelo NSQ, mas independem deste para gerar ritmos, sendo conhecidos como osciladores ou relógios periféricos (MENDOZA, Jorge;

CHALLET, Etienne, 2009, TOSINI, Gianluca, et al., 2008, HARBOUR, 2013) (Figura 01).

Figura 01. Esquema simplificado do sistema de temporização circadiano de mamíferos. Os ritmos biológicos em mamíferos são sincronizados por pistas ambientais, também conhecido pelo termo alemão zeitgebers (doadores de tempo). A principal o principal zeitgeber é a luz, em adição têm-se atividade, temperatura e alimentação. Estes, sincronizam a atividade do núcleo supraquiasmático (NSQ), o relógio central de mamíferos, que imitará sinais extracelulares para os relógios periféricos, no cérebro, como o hipocampo, corpo estriado e cerebelo, e para os outros relógios periféricos situados ao longo do corpo do organismo, como o fígado, coração, rins e etc. Regulando sistemicamente e ritmicamente os diversos eventos bioquímicos, fisiológicos e comportamentais essenciais a sobrevivência dos seres vivos.



Fonte: Adaptada de VIDENOVIC, Aleksandar et al., 2014.

2.2. MECANISMOS MOLECULARES DO RELÓGIO CIRCADIANO EM MAMÍFEROS

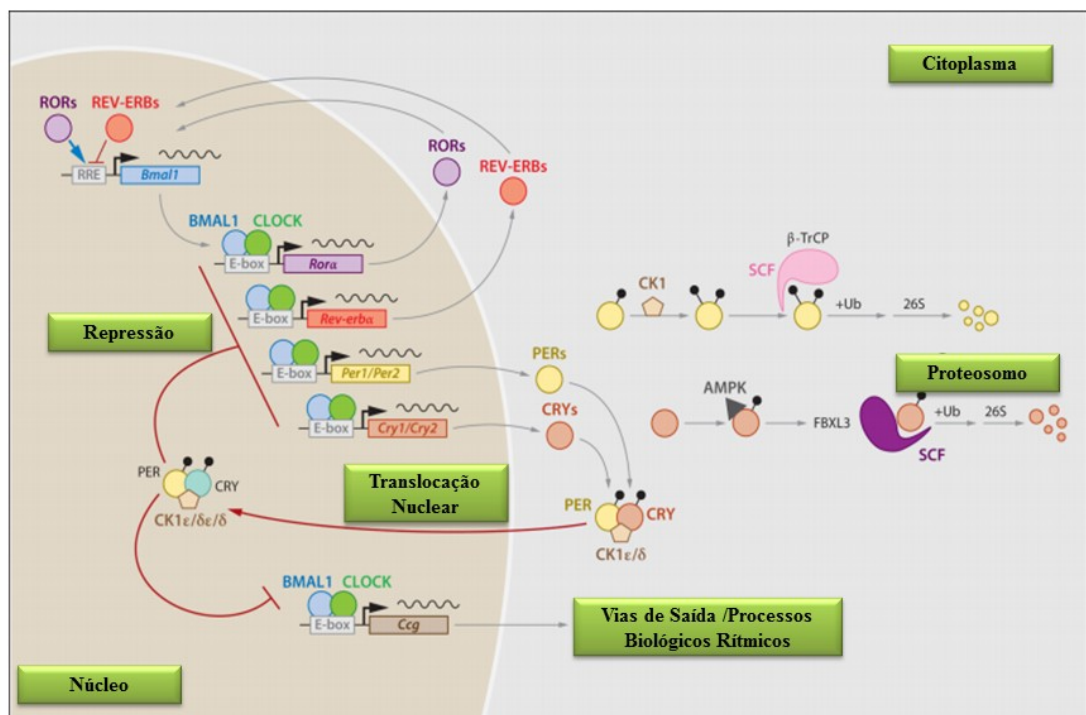
O mecanismo molecular circadiano é constituído, fundamentalmente, por alças de retroalimentação transcricionais e traducionais, que são mantidas por um conjunto de genes altamente conservados entre os animais, conhecidos como genes circadianos (HASTINGS, Michael H.; REDDY, Akhilesh B.; MAYWOOD, Elizabeth S. A., 2003; Lowrey e Takahashi, 2004). Em mamíferos, o mecanismo central do relógio molecular envolve mais de dez genes conhecidos, tais como *Clock* (*Circadian locomotor output cycles kaput*) (VITATERNA, M. Hotz et al.,1994); *Bmal1* (*Brain and muscle-Arnt-like protein-1*), também conhecido na espécie humana como *Arntl* (aryl hydrocarbon receptor nuclear translocator like (BUNGER, Maureen K. et al.,2000); *Per 1* (*Period1*), *Per 2* (*Period2*), *Per 3* (*Period3*) (ZYLKA et al., 1998; MALER, Thomas et al.,1992; TAKUMI et al., 1998); *Cry 1* *Cryptochrome1* e *Cry2* *Cryptochrome 2* (GRIFFIN, Edmund A.; STAKNIS, David; WEITZ, Charles J.,1999; VAN DER HORST et al., 1999); *Nr1d1* nuclear receptor subfamily 1, group D, member 1 e *Nr1d2* nuclear receptor subfamily 1, group D, member 2, também conhecidos como *Rev-erba* e *Rev-erbβ*, respectivamente (PREITNER *et al.*, 2002; PREITNER *et al.*, 2003); *Rora* RAR-related orphan receptor alpha, *Rorb* RAR-related orphan receptor beta, *Rorc* RAR-related orphanreceptor gamma (SATO et al., 2004); *Tim* *Timeless* (KOIKE et al., 2013); os genes *Dec1* também conhecido como *Bhlhe40* basic helix-loop-helix family, member e 40, *Dec2* também conhecido como *Bhlhe41* basic helix-loop-helix family, member e 41 (HONMA et al., 2002) e *Npas2* neuronal PAS domain protein 2 (REICK, Martin et al.,2001).

Uma primeira alça de retroalimentação positiva inclui elementos membros da família de fatores de transcrição do tipo *basic helix-loop-helix*, como *Clock* e *Bmal1*. As proteínas CLOCK e BMAL1 se associam, formando um heterodímero, que se transloca do citosol para o núcleo celular e inicia a transcrição, ligando-se à região regulatória E-box (5'-CACGTG-3') de genes alvos como, *Per1*, *Per2*, *Cry1* e *Cry2*, de outros genes controlados pelo relógio (*Clock Controlled Genes* - CCGs) (kO, Caroline H. e Takahashi, Joseph S., 2007; DUNLAP, Jay C., 1999; PARTCH, Carrie L.; GREEN, Carla B.; TAKAHASHI, Joseph S.,2014; SHEARMAN, Lauren P. et al.,2000;TOSINI, Gianluca et al.,2008). Com o aumento dos níveis de proteínas PER1, PER2, CRY1 e CRY2, formam-se heterodímeros, que reprimem sua transcrição, através da inibição de seu próprio ativador, o complexo CLOCK/BMAL1. A

atuação repressora de PER e CRY consiste, desta forma na segunda alça de retroalimentação, neste caso negativa. (ALBRECHT, Urs; EICHELE, Gregor, 2003; TAKAHASHI, Joseph S. et al, 2008).

Outra alça de retroalimentação é induzida pelo heterodímero CLOCK/BMAL1, atuando na transcrição dos genes *Nr1d1* e *2* e *Rors*, que competem pela ligação ao mesmo sítio no promotor de *Bmal1*. Assim, RORs ativam a transcrição de *Bmal1*, enquanto NR1D1 e *2* reprimem o processo de transcrição de *Bmal1* (AKASHI, Makoto; TAKUMI, Toru, 2005; TRIQUENEAUX, Gérard et al., 2004). A alça molecular de retroalimentação transcricional e traducional em mamíferos está representada na figura 02.

Figura 02. Alça de retroalimentação molecular transcricional e traducional do sistema circadiano em mamíferos. As setas azuis representam a alça de retroalimentação positiva e as vermelhas a alça de retroalimentação negativa. São descritas também vias de degradação e fosforilação de proteínas e a regulação dos CCGs (Genes Controlados pelo Relógio) como via de saída.

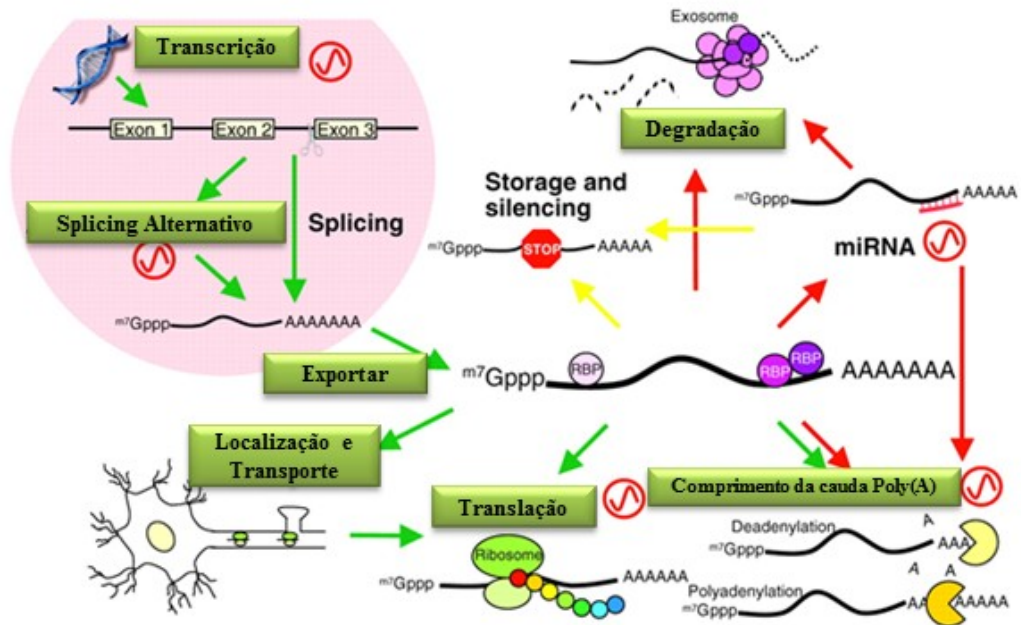


Fonte: Adaptada de MOHAWK, Jennifer A.; GREEN, Carla B.; TAKAHASHI, Joseph S., 2012.

Outros níveis de regulação são importantes para o funcionamento dos relógios circadianos, e envolvem processos como: fosforilação, splicing, degradação, tradução, ubiquitinação, sumoilação, regulação do comprimento da cauda poli (A), atuação de

microRNAs e vias de sinalização celular (KOJIMA, Shihoko; SHINGLE, Danielle L.; GREEN, Carla B.,2011; BECHTEL, William.,2016) (Figura 03)

Figura 03. O envolvimento dos mecanismos pós-transcricionais na regulação dos ritmos circadianos. Os processos destacados pelo símbolo (🕒) indicam que os mesmos possuem regulação conhecida pelos relógios circadianos. As setas verdes, amarelas e vermelhas indicam vias que conduzem aos processos de tradução e silenciamento traducional e degradação de RNAm, respectivamente. São descritos também as RBP (RNA-binding proteins) proteínas de ligação em RNA, e os m7Gppp (7-methylguanosine cap) cap 7 metilguanósina.



Fonte: Adaptada de KOJIMA, Shihoko; SHINGLE, Danielle L.; GREEN, Carla B.,2011.

A nível pós-transcricional as proteínas FBXL3 (F-box and leucine rich repeat protein 3) (SIEPKA, Sandra M. et al.2007) FBXL21 (F-box and leucine rich repeat protein 21) (DARDENTE, Hugues et al.2008), juntamente com o complexo de ubiquitina ligase E3 SCF (*Skp1-Cullin-F-box protein*), da enzima β -TrCP e o complexo proteossomo 26S atuam degradando as proteínas PER e CRY. Outras proteínas como as quinases CK1 ϵ (casein kinase 1, ϵ), CK1 δ (casein kinase 1, delta) e a AMP kinase (AMPK) atuam fosforilando as proteínas PER e CRY (LOWREY e TAKAHASHI, 2004; TAKAHASHI *et al.*, 2008; MOHAWK, Jennifer A.; GREEN, Carla B.; TAKAHASHI, Joseph S.,2012; TAKAHASHI, J. S. 2015). Além destas, outras quinases envolvidas em vias de sinalização celular como a PKA (protein kinase A), CaMK II (calmodulin-dependent kinase) e a MAPK (mitogen-activated protein kinase) estão envolvidas na regulação da ativação da proteína CLOCK (WEBER, Frank et al., 2006).

Compreender a interação entre mecanismos transcricionais e pós-transcricionais é importante para o entendimento do panorama geral da expressão de genes circadianos e seu envolvimento com os ritmos em cada tecido e célula (PARTCH, Carrie L.; GREEN, Carla B.; TAKAHASHI, Joseph S., 2014.)

2.3 RITMOS CIRCADIANOS NO CÉREBRO

2.3.1. O RELÓGIO CENTRAL

O NSQ é considerado o oscilador central em mamíferos, formado por um par de núcleos localizados no sistema nervoso central de vertebrados, na porção anterior do hipotálamo sendo constituídos por, aproximadamente, 10.000 neurônios em camundongos, e 50.000 neurônios em humanos (MARQUES e BARRETO 2003; BUHR e TAKAHASHI, 2013). O NSQ é subdividido em duas regiões anatômicas: uma ventrolateral, chamada de *core*, e outra região dorsomedial e rostral, conhecida como *shell*. A região *core* está localizada acima do quiasma óptico. Ela recebe inervação direta do trato retino hipotalâmico (TRH) e caracteriza-se pela presença de um neuropeptídeo, o polipeptídeo intestinal vasoativo (VIP). Outros neuropeptídeos também estão presentes, mesmo em pequenas quantidades, tais como peptídeo liberador de gastrina (GRP) e calretinina. (WELSH, David K.; Takahashi, Joseph S.; KAY, Steve A., 2010).

A região *shell* é caracterizada pela presença de arginina vasopressina (AVP) e outros neuropeptídeos podem estar presentes, como o met-enkefalina e a angiotensina II (BROOKS, Elisabeth; CANAL, Maria M.; Marques e Barreto 2003; ABRAHAMSON, Eric E.; MOORE, Robert Y., 2001). A retina e o TRH representam a via de entrada de sinais luminosos em mamíferos. (Lowrey e Takahashi, 2004).

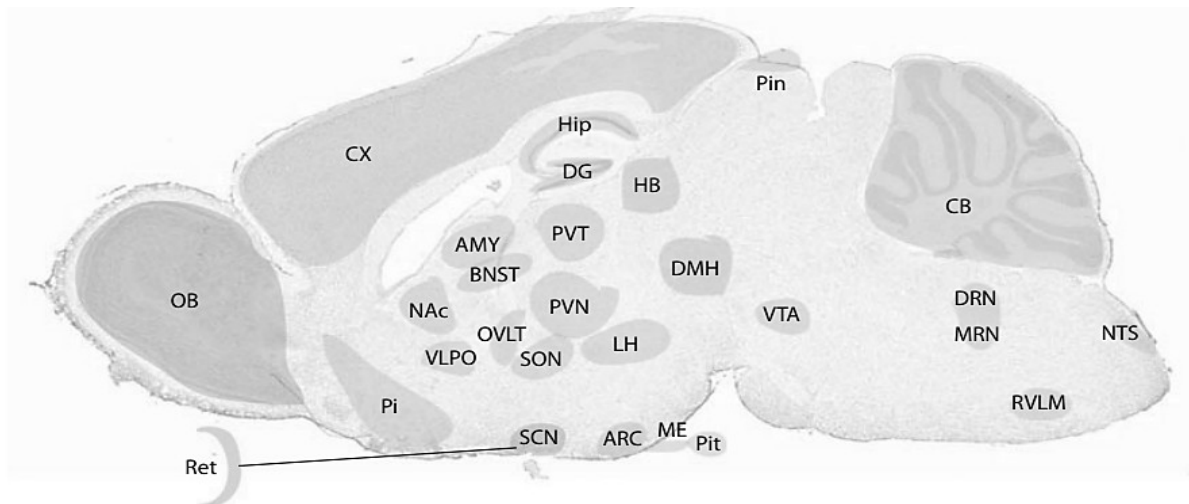
A informação fótica é recebida pela retina por meio de células fotorreceptoras e são transmitidas para as células ganglionares da retina, que em seguida transmite a informação fótica para o cérebro via TRH, cujos axônios projetam-se até o NSQ, constituindo a via ou conexão aferente. Sinais extracelulares partem do NSQ para outras regiões do cérebro, incluindo o prosencéfalo basal, tálamo e outras áreas do hipotálamo, constituindo assim as vias eferentes, que controlam os ritmos circadianos, presentes no organismo. (BROOKS, Elisabeth; CANAL, Maria M., 2013; MARQUES E BARRETO, 2003).

O NSQ sincroniza outros osciladores em todo o cérebro e tecidos periféricos através de diversas vias, incluindo conexões neurais autonômicas e hormônios, que contribuem para sincronizar as células da maioria dos tecidos. Por exemplo, a regulação da expressão de genes circadianos no fígado é impulsionada, por sinais emanados do NSQ (Figura 1) (WELSH, David K.; TAKAHASHI, Joseph S.; KAY, Steve A. 2010; GUILDING, Clare; PIGGINS, Hugh D., 2007; DAMIOLA, Francesca et al., 2000; YAMAZAKI, Shin, et al., 2000).

2.3.2. RELÓGIOS PERIFÉRICOS

Em 2004, o bulbo olfatório foi identificado como o primeiro relógio periférico no cérebro que possui oscilação circadiana independente do NSQ em mamíferos. A taxa de neurônios disparados por células do bulbo olfatório cultivadas *in vitro* demonstrou oscilação circadiana por um período aproximado de 24 horas (GRANADOS- FUENTES, Daniel et al., 2004; HERZOG, Erik D.,2007). Regiões individuais do cérebro, tais como o hipotálamo, o corpo estriado e os córtices contêm sua própria oscilação circadiana, independentemente do NSQ e são capazes de gerar ritmos circadianos quando são isolados dos organismos e cultivados *in vitro*. (ABE et al., 2002; HERZOG, Erik D.,2007). O cerebelo também foi descrito como um relógio periférico, pois é comprovada sua associação no controle do ritmo circadiano alimentar (MENDOZA, Jorge et al.,2010). Além disso, a taxa de disparo neuronal das células *Purkinje*, presentes no cerebelo, demonstrou oscilação durante aproximadamente três dias. (MORDEL, Jérôme et al.2013). A figura 04 mostra uma representação dos potenciais osciladores periféricos no cérebro de mamíferos e o oscilador central.

Figura 04. Potenciais osciladores periféricos no cérebro de mamíferos e o oscilador central. AMY, amígdala; ARC, núcleo arqueado; BNST, núcleo da estria terminal; CB, cerebelo; CX, córtex; DG, giro denteado; DMH, hipotálamo dorso medial; DRN, núcleo dorsal da rafe; HB, habenula; HIP, hipocampo; LH, hipotálamo lateral; ME, eminência mediana; MRN, núcleo mediano da rafe; NAc, núcleos accumbens; NTS, núcleo do trato solitário; OB, bulbo olfatório; OVLT, órgão vascular da lâmina terminal; Pi, córtex piriforme; Pin, glândula pineal; Pit, glândula pituitária; PVN, núcleo paraventricular do Hipotálamo; PVT, núcleo paraventricular do tálamo; Ret, retina; RVLM, medula rostral ventrolateral; SCN, núcleo supraquiasmático; SON, núcleo supra-óptico; VLPO, área pré-óptica ventrolateral; VTA, área tegumental ventral.



Fonte: DIBNER, Charna; SCHIBLER, Ueli; ALBRECHT, Urs.,2010.

A descoberta de genes circadianos possibilitou a identificação de áreas do cérebro que possuem a maquinaria molecular necessária para a geração dos ritmos biológicos. Assim, a expressão rítmica de genes circadianos foi identificada numa série de regiões do cérebro incluindo o tálamo e hipotálamo, bulbo olfatório e o cerebelo (FEILLET, Céline A. et al.,2008). Os genes *Clock* e *Bmal1* apresentaram expressões circadianas relatadas no córtex piriforme, bulbo olfatório, hipocampo, corpo estriado e cerebelo de ratos (NAMIHIRA, Masakazu et al.,1999).

Foi demonstrado que os RNAm e as proteínas dos genes *Pers* são ritmicamente expressas em culturas de células neurais do hipocampo, corpo estriado, córtex piriforme e cerebelo de rato (ABE et al., 2002). *Dec1* e *Dec2* também possuem expressão circadiana comprovada em cultura de células do córtex frontal (ROSSNER, Moritz J. et al.,2008). *Npas2* é expresso em algumas regiões do cérebro como, no hipocampo, corpo estriado e córtex. Camundongos *knockout* para esse gene apresentaram alterações nos ritmos circadianos de atividade e repouso, e no ritmo alimentar. Estas alterações indicam que a expressão de *Npas2*

nas regiões citadas está envolvida com o comportamento em camundongos (DUDLEY, Carol A. et al.,2003). Além disso, os genes *Per1*, *Per2*, *Per3*, *Cry1*, *Bmal1*, *Nr1d1* e *Dbp* possuem expressão circadiana comprovada em cerebelo de ratos (RATH, Martin F.; ROHDE, Kristian; MØLLER, Morten).

2.4 IDENTIFICAÇÃO DE GENES CIRCADIANOS E CCGs

Diferentes abordagens e métodos foram utilizados para a identificação dos genes circadianos e de seu papel na regulação dos ritmos circadianos em mamíferos. A identificação dos principais genes circadianos iniciou-se a partir de 1990. Em 1994 Martha Vitaterna e colaboradores, identificaram o primeiro gene relógio envolvido com o sistema circadiano de mamíferos. Após realizarem uma análise de mutagênese em camundongos, observaram que os animais apresentavam mudanças no padrão do comportamento de atividade e repouso. Posteriormente, encontraram que um gene estava envolvido com comportamento circadiano de atividade e repouso, localizado no cromossomo 5 de camundongos, equivalente ao cromossomo 4 em humanos, o qual foi denominado de *Clock* (VITATERNA, M. Hotz et al.,1994).

Em 1992, Maler e colaboradores avaliaram que sequências conservadas do gene *Per1* em drosófila eram também expressas em NSQ de mamíferos (MALER, Thomas et al.,1992). Cinco anos depois, em 1997, complementando esse estudo, TEI e colaboradores 2007 confirmaram esse dado através da análise de pequenas sequências do gene por IMS-PCR (*intra-module scanning-polymerase chain reaction*), e confirmaram através da técnica de hibridização *in situ* (*HIS*) a expressão circadiana deste gene em NSQ de camundongos (TEI, Hajime et al., 1997). Logo em seguida, em 1998, Takumi e colaboradores também demonstraram a expressão circadiana do gene *Per2* em NSQ de camundongos, através da técnica de *HIS* (TAKUMI, Toru et al., 1998). Nesse mesmo ano, Mark Zylka e colaboradores demonstraram a expressão circadiana de *Per3* também em NSQ de camundongos (ZYLKA, Mark J. et al., 1998).

Em 1999, VAN e colaboradores, mostraram que as proteínas CRY1 e CRY2 são essenciais para a manutenção dos ritmos circadianos. Por meio de análises de comportamento de atividade repouso em camundongos e da quantificação de RNA e proteínas pelas técnicas de *RT-PCR* e *Southern blot*, respectivamente, eles encontraram que camundongos *knockout* para o gene *Cry2* apresentaram alongamento do período endógeno em condições de escuro

constante, e que animais *knockout* para o gene *Cry1* apresentaram encurtamento de seus períodos endógenos. Quando os animais são *knockout* para ambos os genes, eles apresentam arritmia no ritmo de atividade e repouso (VAN DER HORST, Gijsbertus TJ et al.,1999). *Bmall*, outro importante gene relógio teve seu envolvimento comprovado com os ritmos circadianos no ano de 2000, através de análises comportamentais e de expressão gênica por *hibridização in situ*. Assim, BUNGER e colaboradores demonstraram que camundongos *knockout* para este gene, apresentaram arritmia no perfil comportamental de atividade e repouso em condições de escuro constante (BUNGER, Maureen K. et al.,2000).

Há um constante interesse em descobrir novos genes circadianos e *CCGs*, com o objetivo de entender o funcionamento desses ritmos em mamíferos. Com o advento de técnicas de análises globais como, RNA-seq, *Microarray*, *ChiP-seq* (*Chromatin immunoprecipitation sequencing*) e *screening* por *RNAi* (RNA de interferência), dentre outras, têm-se acelerado a capacidade de encontrar esses novos membros do sistema circadiano (BHARGAVA, Anuprabha; HERZEL, Hanspeter; ANANTHASUBRAMANIAM, Bharath, 2015; OJIMA, Shihoko; GREEN, Carla B.,2014; HOGENESCH, John B.; UEDA, Hiroki R.,2011). Um estudo em larga escala utilizando *RNAi* revelou que a casein kinase 2 (CK2) está envolvida com o relógio circadiano de mamíferos, atuando na fosforilação da proteína PER2 (MAIER, Bert et al.,2000).

Análises de *ChiP-seq* para sete fatores de transcrição circadianos, BMAL1, CLOCK, NPAS, CRY1, CRY2, PER1 e PER2 identificaram que centenas de genes possuem sítios de ligação para estes fatores (KOIKE et al., 2013). FANG e colaboradores (2014) encontraram, através de análises de *ChiP-seq* em algumas fases do dia, em fígado de camundongo *Knockout*, que diversos genes possuem sítios de ligação para o fator de transcrição NR1D1, tendo seus transcritos inibidos ou ativados após a deleção do mesmo.

Um estudo baseado em uma abordagem genômica identificou que a proteína RORa é um componente do relógio circadiano de mamíferos, atuando com um ativador do gene *Bmall* (SATO, Trey K. et al.,2004). Não somente genes circadianos têm sido identificados. Há uma pesquisa intensa para identificar *CCGs* que possam estar associados com diferentes fenótipos circadianos. Uma análise de transcriptoma circadiano em 12 estruturas e tecidos de camundongos, revelou que até 43 % dos transcritos em determinados tecidos ou estrutura possuem expressão circadiana (ZHANG, et al.,2014). Outro estudo ainda com transcriptoma em 6 estruturas do cérebro de pessoas saudáveis considerando a hora do óbito, revelou que

mais de 100 transcritos comuns apresentaram expressão circadiana nas estruturas avaliadas. No total, 3.573 genes apresentaram variação temporal no cérebro humano, considerando todas as regiões analisadas. (LI, J. Z. et al.,2013).

MARET e colaboradores (2007) realizou análises do perfil circadiano do transcriptoma no cérebro e fígado de camundongos privados de sono. Eles encontraram que 391 transcritos permanecem rítmicos em camundongos privados de sono (grupo experimental) e para seu grupo controle demonstraram que 2.032 transcritos possuem expressão circadiana. Encontrando uma diminuição da quantidade de transcritos que oscilam em animais que passaram por privação de sono.

HUGHES e colaboradores (2009) realizaram uma comparação do perfil harmônico circadiano de transcriptoma do fígado de camundongos, e de duas linhagens celulares (NIH3T3 e U2OS), e demonstraram através de dois métodos estatísticos (*Fisher's G-test* e *COSOPT*), considerando o FDR (*false discovery rate*) $<0,05$, que 3.667 transcritos no fígado possuem ritmo de expressão em torno de 24 horas, menos de 12 transcritos nas linhagens celulares, possuem ritmo de expressão em torno de 24 horas. Além disso, 260 transcritos apresentaram ritmos ultradianos em torno de 12 horas e 63 em torno de 8 horas, harmônicos dos transcritos que apresentaram 24 horas. Os autores discutem diante do critério estatístico estabelecido apenas essa quantidade oscilaram e que talvez existam outros transcritos que oscilem circadianamente nas linhagens celulares, pois em cultura celular há uma perda de ritmicidade dos transcritos.

Estudos com análise em larga de proteoma circadiano demonstraram que em fígado de camundongo 20% das proteínas apresentaram ritmicidade circadiana (REDDY, A. B. et al., 2006). Contrariamente, em outro estudo, 6 % de 3.000 proteínas avaliadas apresentaram ritmicidade circadiana em fígado de camundongos (ROBLES, Maria S.; COX, Jürgen; MANN, Matthias.,2014). Em outro estudo semelhante, também em fígado de camundongo, das 5.000 proteínas,195 (~5%) apresentaram variação circadiana (MAUVOISIN et al., 2014). Uma análise com 2112 proteínas em NSQ de camundongos demonstrou que 20% apresentaram ritmicidade circadiana (CHIANG et al., 2014).

No entanto, apesar destas abordagens em larga escala terem sido amplamente utilizadas na identificação de moléculas funcionais envolvidas no relógio circadiano mamífero, elas

apresentam algumas limitações. Os resultados de estudos com transcriptoma muitas vezes são inconsistentes, pois, em muitos estudos, não é claro quais genes tem papéis importantes para os ritmos. Mesmo no caso de estudos que utilizam a mesma abordagem metodológica e avaliação no mesmo tecido, apresentam incompatibilidade entre os resultados. Essas inconsistências também abrangem outros estudos com análises em larga escala como RNA-seq, proteoma, Microarray e Nascent-seq. Porém, podem ser derivadas do tipo de algoritmo utilizado para detectar ritmicidade, pois há diversos com essa função, como *JTK_CYCLE*, *ARSER*, *COSOPT*, *Fisher's G test*, e *CircWave* e também pela normalização de tecido por exemplo, isso interfere na detecção de genes que apresentam baixas amplitudes de expressão (KOJIMA, Shihoko; GREEN, Carla B, 2014).

2.5 MINERAÇÃO DE DADOS NA BUSCA DE NOVOS GENES CANDIDATOS AO CONTROLE DOS RITMOS CIRCADIANOS EM MAMIFEROS

Abordagens mais recentes como a mineração dados, definida como uma análise de grandes conjuntos de dados com o objetivo de encontrar novas relações entre eles, resumindo-os em forma compreensíveis e úteis para gerar novos conhecimentos (HAND, David J.; MANNILA, Heikki; SMYTH, Padhraic., 2001), é também utilizada em estudos com avaliação rítmica da expressão de genes e em estudos cronobiológicos com humanos, roedores e plantas (SUKUMARAN, Siddharth et al. 2010; FILICHKIN, Sergei A. et al. 2011; ALMON, Richard R. et al. 2008; LOPES, Robson da Silva et al. 2013). O gene *Chrono*, por exemplo, foi identificado como um importante componente do relógio circadiano em mamíferos, através de uma abordagem computacional. ANAFI e colaboradores 2014 utilizaram uma abordagem probabilística computacional para integrar informações de bancos de dados heterogêneos com informações de larga escala de transcriptoma circadiano, expressão gênica em diferentes tecidos, filogenia de genes, vias de interações de proteínas e RNAi, para identificar genes circadianos candidatos que apresentaram evidências de envolvimento com os ritmos circadianos de acordo com esses dados de larga escala e selecionaram aproximadamente 1000 genes mais bem ranqueados. Posteriormente, eles selecionaram os 25 primeiros genes mais ranqueados dentre os 1000 selecionados e avaliaram *in vitro* a potencial interação física destes com as principais proteínas relógio. Desta forma, eles encontraram que o gene *Chrono* interage com as proteínas BMA11, interrompendo sua atividade transcricional. Além disso, camundongos *Knockout* para este gene apresentaram

ritmos prolongados de atividade locomotora. Desta forma, comprovaram o envolvimento desse gene com os ritmos circadianos de mamíferos.

BHARGAVA e colaboradores (2015), selecionaram 1.000 genes candidatos a reguladores dos ritmos circadianos, obtidos do estudo ANAFI e colaboradores (2014) realizaram um estudo de mineração de dados com outros estudos circadianos de proteoma, *ChiP-seq* e interação proteína-proteína, demonstrando 11 genes candidatos, que podem fazer parte do sistema molecular de temporização de mamíferos, propondo seis destes genes que estão envolvidos com metabolismo de câncer, para investigações experimentais.

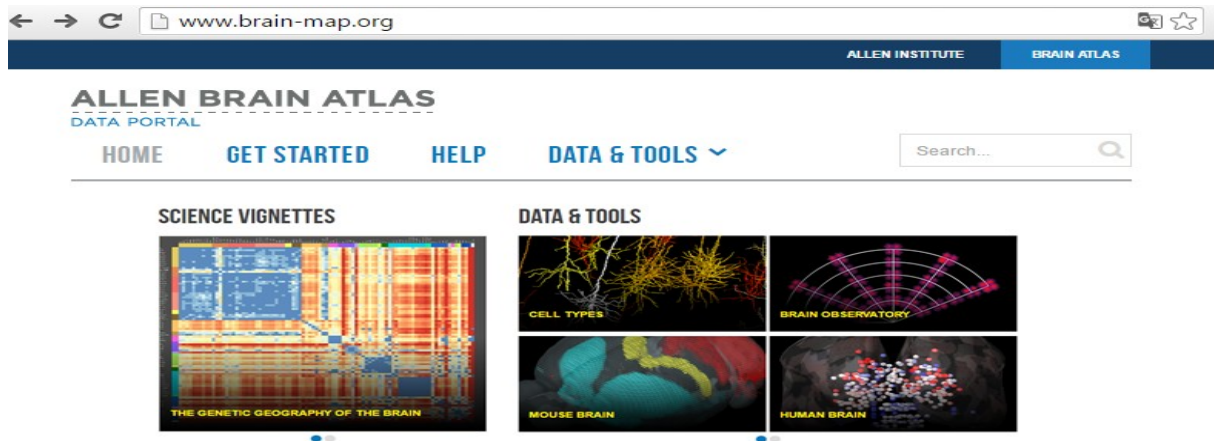
LEHMANN e colaboradores 2015, com o objetivo de encontrar genes candidatos a regulação circadiana de processos celulares, desenvolveram um *pipeline* de bioinformática que integra diferentes fontes e tipos de informações de mineração de dados coletados no portal *Pubmed*, além de análises de co-expressão de genes. Com esta estratégia, identificaram 118 potenciais genes controlados pelo relógio e que estão envolvidos com a regulação circadiana dos ritmos em mamíferos. Os resultados foram confrontados com os estudos que HUGHES e colaboradores (2009) realizaram análises em larga escala com estudos de transcriptoma circadiano de camundongos e cultura de células, respectivamente, mostrando que os 118 genes potenciais exibiram expressão circadiana nestes trabalhos. NAYAK, e colaboradores, 2009, por meio das análises nos bancos de dados de *ChiP-seq* para as proteínas BMAL1, RORa e b e Nr1d1 e 2, identificaram que 59 genes possuem sítios alvos para as proteínas Nr1d1 e 2, 19 para as proteínas RORa e b e dois para a proteína BMAL1 e que podem estar envolvidos com os ritmos circadianos.

2.6 ALLEN BRAIN ATLAS

A partir da fundação do *Allen Institute for Brain Science* em 2003, pelo filantropo *Paul G. Allen*, cuja a finalidade era realizar pesquisas voltadas para a investigação de como os cérebros de humanos saudáveis ou com doenças funcionam, *Paul Allen* resolve tornar os dados oriundos destes estudos, globais e acessíveis a toda população. Desta forma, ele e uma grande equipe de cientistas de diferentes áreas criou em 2004 O *Allen Brain Atlas* (ABA), um portal *on line* (<http://www.brain-map.org/>) (Figura 05) que reúne dados públicos em pesquisas no cérebro de humanos, camundongos e primatas, com enfoque em estudos de expressão gênica em larga escala, conectividade neurais, caracterização celular e neuroanatomia, distribuídos em 11 diferentes bancos de dados (Tabela 01) (<http://www.brain->

[map.org/](http://www.brain-map.org/); JONES, Allan R.; OVERLY, Caroline C.; SUNKIN, Susan M.,2009; SUNKIN, Susan M. et al.,2012; KUAN, Leonard et al., 2015).

Figura 05. Página principal do Portal *Allen Brain Atlas*. Os usuários podem acessar e pesquisar todos os conjuntos de dados nas diferentes ferramentas, escolhendo a opção (*DATA E TOOLS*). Para cada conjunto de dados distribuídos nas diferentes ferramentas, o portal disponibiliza na opção (*GET STARDED*) tutoriais de uso explicativos em forma de vídeos, visão geral de cada ferramenta e dados principais. Além disso, o portal disponibiliza ajuda *on line* na opção (*HELP*).



Fonte: (<http://www.brain-map.org/>)

Tabela 01. Os principais recursos públicos do portal *Allen Brain Atlas*, com suas respectivas características e ordenados de forma cronológica de desenvolvimento.

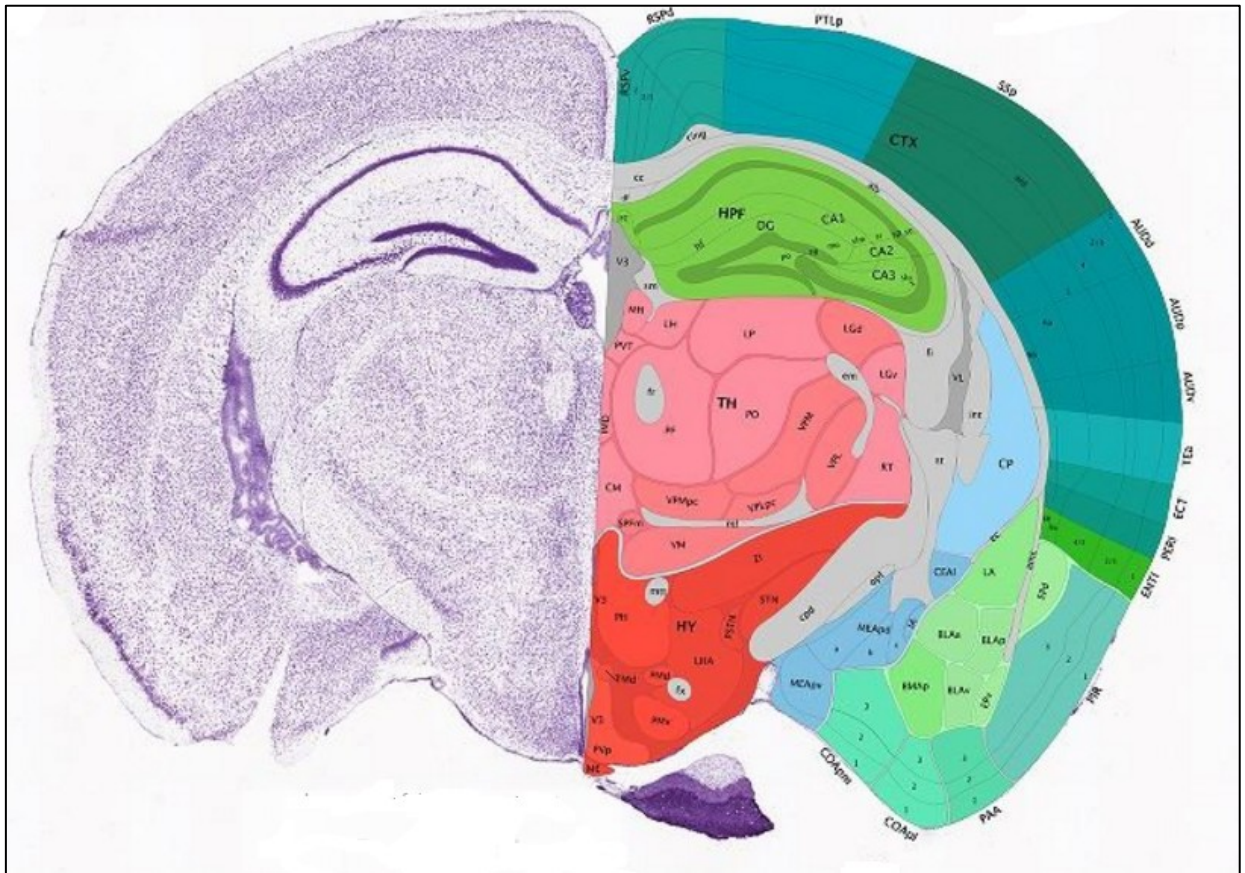
Bancos de dados	Características principais
Allen Mouse Brain Atlas	Primeiro recurso disponível no portal <i>Allen Brain Atlas</i> , foi criado em 2004, disponibilizando dados de expressão gênica espacialmente mapeado e quantificado através da técnica de <i>HIS</i> de todo o cérebro de camundongos.
Allen Mouse Spinal Cord Atlas	É um recurso abrangente criado em julho de 2008, que detalha a expressão de genes na medula espinal, incluindo imagens baseadas na técnica de <i>HIS</i> , com alta resolução microscópica a nível celular e anatômica de toda a medula espinal de camundongos jovens e adultos.
Allen Developing Mouse Brain Atlas	É um atlas detalhado que explora dados do período e estrutura em que aproximadamente 2.000 genes são ativados durante o desenvolvimento cérebro de camundongos, desde o período embrionário até a velhice. Foi criado em novembro de 2008.
Non-Human Primate Atlas	Criado em 2009, é um atlas neuroanatômico que explora a arquitetura molecular e celular do desenvolvimento do cérebro de primatas pós-natal, com relevância direta para entender o desenvolvimento do cérebro humano.
Allen Human Brain Atlas	É um atlas multimodal de expressão genica do cérebro de humanos. Reunindo dados anatômicas de todas as estruturas cerebrais por meio de imagens de ressonância magnéticas, por tensores de difusão e histológicas, e genômicos através da expressão de todos os genes utilizando as técnicas de Microarray e <i>HIS</i> . Este atlas foi desenvolvido em 2010.
Brain Span Atlas of the Developing Human Brain	É um atlas neuroanatômico e de expressão de genes baseada nas técnicas de <i>Microarray</i> , <i>HIS</i> e RNA-seq do cérebro de humanos em desenvolvimento. Este recurso foi desenvolvido no ano de 2010.
Allen Mouse Brain Connectivity Atlas	É um atlas desenvolvido em 2011, de alta resolução de conexões neurais do cérebro de camundongo. Inclui imagens de projeções axonais dos cérebros de camundongos transgênicos, caracterizando a expressão de Cre e de outros genes. Estes dados também são apresentados em 3D através do software Brain explorer.
Allen Cell Types	É recurso de caracterização funcional e estrutural dos tipos celulares do cérebro de camundongos adultos. A partir de dados eletrofisiológicos e morfológicos de células individuais. Foi criado em 2016.

Glioblastoma Atlas	É recurso que explora a base anatômica e genética a nível celular e molecular de tumores cerebrais (Glioblastoma). Desenvolvido em 2015.
Aging, Dementia and Traumatic Brain Injury Study	É um recurso detalhado da caracterização neuropatológica, molecular e do transcriptoma de cérebros de humanos envelhecidos, com demência, expostos e não expostos a traumatismo craniano. Foi criado em 2016.
Allen Brain Observatory	É recurso criado em 2016, que mensura de forma sistemática as repostas de neurônios em todas as camadas e áreas do córtex visual da linhagem de camundongos transgênicos para o gene Cre.

2.6.1 ALLEN MOUSE BRAIN ATLAS

O *Allen Mouse Brain Atlas* (<http://mouse.brain-map.org/>) foi o primeiro recurso do portal *Allen Brain Atlas*, criado em dezembro de 2004. Consiste em um mapa alta resolução tridimensional de escala genômica, que oferece dados de expressão gênica espacial, de aproximadamente 20.000 genes, em todo o cérebro de camundongos C57BL/6J adultos. O atlas apresenta por meio de imagens nos plano anatômico sagitais de 21 cortes com 200 µm de espessura de 71 estruturas cerebrais e do plano coronal 132 cortes com espessura de 100 µm das diversas estruturas do cérebro, a expressão de genes a nível celular localmente marcado através da técnica de *HIS*, integrando a esta, diversas ferramentas na busca e visualização de dados de genes com padrões de expressão espaciais correlacionadas e um atlas de referência anatômica (figura 06) (<http://mouse.brain-map.org/>; NG, Lydia et al., 2007; NG, Lydia et al., 2009; LEIN, Ed S. et al., 2007).

Figura 06. Expressão do gene relógio *Per1* em diversas regiões do cérebro de camundongos. A expressão foi obtida através da técnica de *HIS* de um corte coronal, hemisfério esquerdo, e no hemisfério direito é apresentado o atlas de referência anatômica.

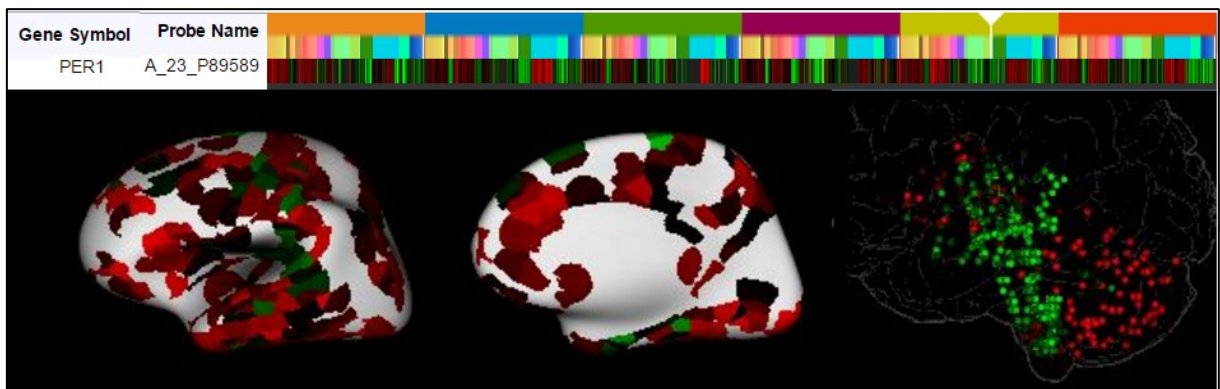


Fonte: (<http://www.brain-map.org/>)

2.6.2 ALLEN HUMAN BRAIN ATLAS

O *Allen Human Brain Atlas* (<http://human.brain-map.org/>) é um atlas multimodal que mapeia a expressão espacial de genes no cérebro de humanos adultos. Os dados de expressão de genes em larga escala nas regiões do córtex cerebral, subcortex, cerebelo, e tronco encefálico de múltiplos cérebros de humanos adultos mortos, foram obtidos pela técnica de Microarray. É possível acessar esses dados pela busca de genes e estruturas anatômicas. Foram coletadas entre 400 e 500 tecidos por hemisférios de cada cérebro, com aproximadamente 90.000 amostras e 60.000 sondas por amostras. Assim, através de um mapa quente (*Heat map*) com escala de cor dada em z score ou \log_2 , é possível visualizar simultaneamente os valores de expressão espacial de múltiplos genes em todas as estruturas do cérebro (Figura 07) (<http://human.brain-map.org/>; SHEN, Elaine H.; OVERLY, Caroline C.; JONES, Allan R.,2012; HAWRYLYCZ, Michael J. et al.,2012).

Figura 07. Mapa quente demonstrando a expressão espacial do gene *Per1* no cérebro de humanos. As imagens correspondem a expressão deste gene no córtex esquerdo, direito e subcortex do cérebro de humano. As regiões cerebrais identificadas pela cor vermelha indica alta expressão de *Per1*, verde baixa expressão.



Fonte: (<http://www.brain-map.org/>)

Além disso, através da técnica *HIS*, é possível acessar dados de expressão e imagens anatômicas de conjunto genes em cinco estudos independentes, em regiões específicas do cérebro. No primeiro estudo, é possível acessar os dados de expressão e imagens anatômicas de 1.000 genes na região visual e temporal do córtex cerebral, o segundo apresenta dados de 67 genes nas áreas subcorticais, o terceiro faz um estudo com os neurotransmissores de 176 genes no córtex e 88 genes no subcortex. No quarto e quinto estudo, são apresentados respectivamente, dados de expressão e imagens anatômicas de 25 genes envolvido com

autismo e de 60 genes envolvidos com esquizofrenia, com caso controles (SHEN, Elaine H.; OVERLY, Caroline C.; JONES, Allan R.,2012; HAWRYLYCZ, Michael J. et al.,2012).

2.6.3 EXPRESSÃO ESPACIAL DE GENES CORRELACIONADOS

A expressão de cada gene nas diferentes estruturas do cérebro de camundongos e humanos foi quantificada pelas técnicas de *HIS* e *Microarray*, respectivamente, sendo representada no cérebro em 3D, por uma rede de pixels, uma estrutura computacional quadriculada na forma de cubos. Cada ponto corresponde a um conjunto de expressão de vários genes. A partir disto, é possível buscar por meio da ferramenta *NEUROBLAST*, disponível no portal, genes com perfil espacial de expressão similar ou oposta a um dado gene de interesse no cérebro como um todo, ou por estruturas específicas, através de uma análise estatística de correlação (coeficiente r de Pearson) (SUNKIN, Susan M. et al.2013). Encontrar genes com um padrão de expressão espacial semelhante a determinado gene conhecido, pode revelar uma relação potencial desconhecida, que envolve esses genes em processos e vias semelhantes (NG, Lydia et al., 2007).

Assim, genes com padrão de expressão espacial correlacionada a genes que apresentam oscilação circadiana ou que são importantes para a geração de ritmos biológicos, são, potencialmente, candidatos a genes circadianos ou *CCGs* com oscilação temporal de expressão naquele tecido. Este tipo de abordagem pode, portanto, possibilitar a identificação de novos genes importantes para o controle do ritmo circadiano.

3.0 JUSTIFICATIVA

A identificação de novos genes circadianos ainda é uma necessidade de pesquisa, tendo em vista a alta complexidade e os diferentes níveis de regulação do sistema molecular em mamíferos. Recentemente, com o acúmulo de dados experimentais de expressão gênica em larga escala, estratégias de mineração de dados tem sido bastante utilizada com esta finalidade. De fato, novos genes circadianos, validados experimentalmente, têm sido descritos a partir deste tipo de abordagem metodológica (ANAFI, Ron C. et al.,2014; BHARGAVA, Anuprabha; HERZEL, Hanspeter; ANANTHASUBRAMANIAM, Bharath.,2015; LOPES, Robson da Silva et al.,2013).

Desta forma, genes ainda não caracterizados como circadianos, que possuem perfis de expressão semelhantes aos principais genes circadianos, em diferentes células ou tecidos, podem:

- 1- Ser um gene circadiano;
- 2- Ser um gene controlado pelo relógio (*clock controlled gene*).

4. OBJETIVOS

4.1 OBJETIVO GERAL

Identificar novos genes candidatos a reguladores do ritmo circadiano no cérebro de mamíferos.

4.2 OBJETIVOS ESPECÍFICOS

- Identificar genes com padrão de expressão espacial correlacionada a genes circadianos no cerebelo de *Mus musculus* e *Homo sapiens*, utilizando dados disponíveis no portal *Allen Brain Atlas*.
- Identificar na literatura e em outras bases de dados se os genes identificados possuem evidências de expressão circadiana e/ou controle do ritmo circadiano.
- Identificar genes candidatos a moduladores do sistema de temporização circadiano de mamíferos.

5. MATERIAL E MÉTODOS

5.1 MINERAÇÃO DE DADOS

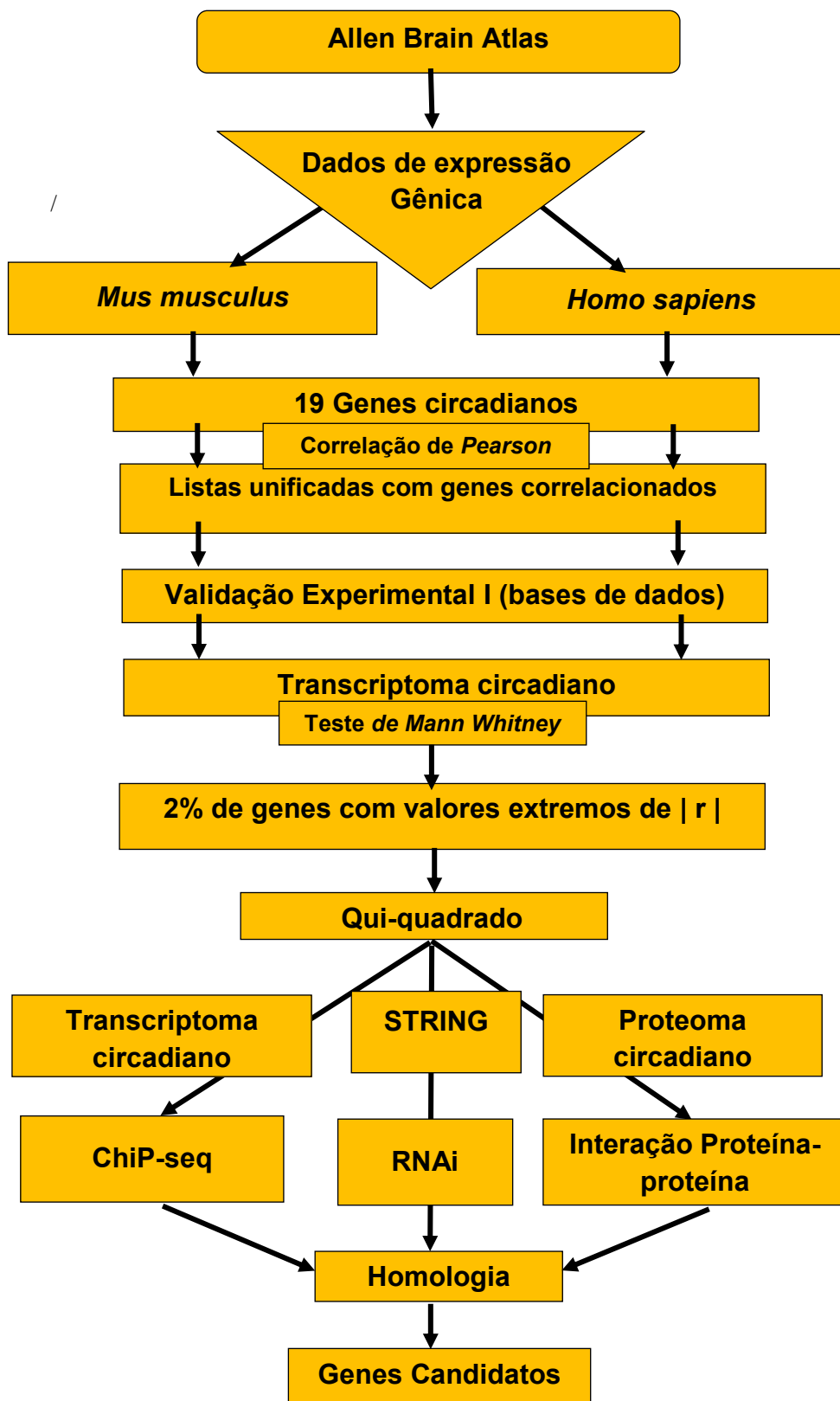
5.1.1 SELEÇÃO DE GENES CORRELACIONADOS AO RITMO CIRCADIANO NO CEREBELO HUMANO E DE CAMUNDONGO

A figura 08 ilustra o delineamento da estratégia geral de mineração de dados para seleção de genes candidatos a serem explorados em ensaios experimentais subsequentes, a partir das informações sobre expressão gênica espacial em larga escala disponíveis no portal *Allen Brain Atlas*. Inicialmente foram coletados no período de janeiro à dezembro de 2016, nas ferramentas “*mouse brain*” e “*human Brain*”, dados sobre genes que possuem correlação de expressão espacial no cerebelo, com 19 genes circadianos (*Bmal1*, *Bmal2*, *Clock*, *Cry1*, *Cry2*, *Csnk1a*, *Csnk1d*, *Csnk1e*, *Fblx3*, *Fblx21*, *Npas2*, *Nr1d1*, *Nr1d2*, *Per1*, *Per2*, *Per3*, *Rora*, *Rorb*, e *Rorc*), das espécies *M. musculus* e *H. sapiens*, respectivamente. Estes 19 genes estão entre os componentes principais do sistema circadiano de mamíferos. Consistem em fatores de transcrição, e proteínas de fosforilação e ubiquitinação envolvidos no controle transcricional e pós-transcricional/traducional oscilatório e de diversos genes circadianos e CCGs. Os dados coletados de cada gene correlacionado foram: nome e símbolo oficial, ID do gene, número do experimento para cada gene, as sondas utilizadas por experimento, a orientação da sonda (sense ou antisense), o plano do corte anatômico histológico por experimento (sagital ou coronal), e o coeficiente de correlação de *Pearson* para cada gene relógio.

O cerebelo foi escolhido com base em alguns fundamentos: **1.** Esta estrutura é descrita na literatura como um relógio circadiano periférico, envolvida com o ritmo alimentar de camundongos, possuindo também expressão rítmica dos genes relógio: *Per1*, *Per2*, *Per3*, *Bmal1*, *Cry* e *Nr1d* (MENDOZA, Jorge et al.,2010; RATH, Martin F.; ROHDE, Kristian; MOLLER, Morten.,2012); **2.** O *Allen Brain Atlas* disponibiliza dados de correlação espacial específicos para esta estrutura, em humanos e camundongos, o que permite uma análise comparativa entre espécies de uma mesma estrutura específica; **3.** Existem dados disponíveis de transcriptoma circadiano para esta região em humanos (LI. J. Z. et al.,2013) e camundongos (ZANGH et al.,2014) o que permite uma comparação em larga escala entre

dados de correlação espacial (*Allen Brain Atlas*) e temporal (estudos circadianos publicados, de transcriptoma, proteoma, chiP-seq, interação proteína-proteína e RNA de interferência).

Figura 08. Fluxograma simplificado das etapas para identificação de genes candidatos ao controle dos ritmos circadianos em mamíferos.



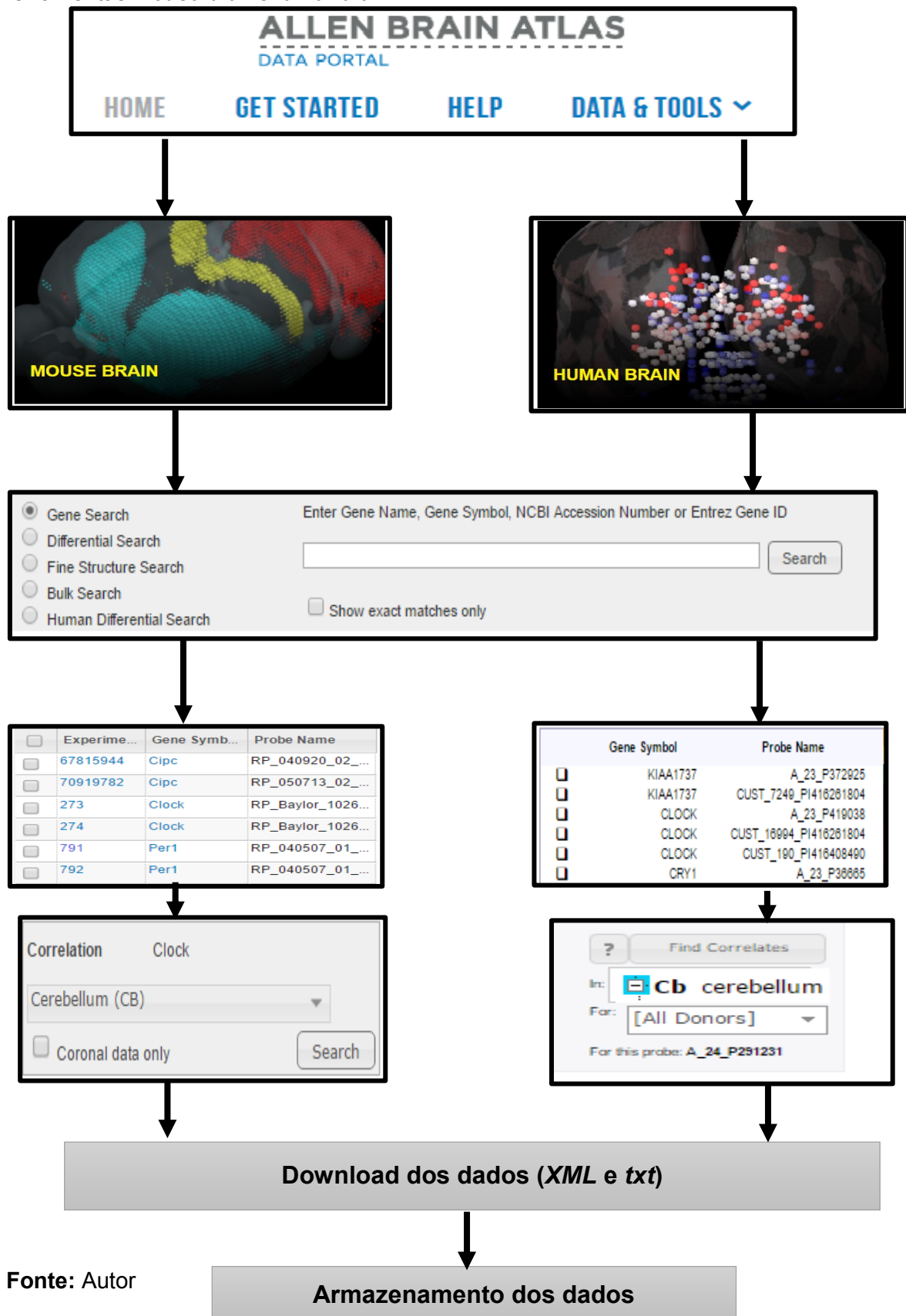
Fonte: Autor

Foi selecionado individualmente cada experimento disponível para cada um dos genes circadianos, gerando assim uma lista de genes com seus respectivos valores de r (coeficiente de correlação *de Pearson*).

Estas listas foram obtidas por *download* no formato de arquivo *XML* e *txt* e armazenadas para posteriores análises. Este processo foi feito separadamente para as espécies *M. musculus* e *H. sapiens* conforme ilustrado na figura 09.

Tendo em vista que um mesmo gene pode estar correlacionado com diferentes genes circadianos, realizamos uma nova triagem para selecionar o maior valor de correlação de um dado gene com qualquer gene relógio analisado, gerando assim um painel de genes correlacionados com o “ritmo circadiano” no cerebelo humano e murino. Desta forma, foi gerada uma única lista de genes por espécie, a qual foi denominada de “*lista mestre*”. Além disto, foi criada uma terceira lista unificada com genes homólogos entre as espécies *M. musculus* e *H. sapiens*, a partir da análise realizada da intersecção dos genes presentes nas duas listas mestres.

Figura 09. Fluxograma ilustrando as etapas seguidas na busca e armazenamento dos dados de genes que possuem expressão espacial correlacionada com 16 genes relógio, em cerebelo das espécies *M. musculus* e *H. sapiens*, no portal *Allen Brain Atlas*, nas ferramentas *mouse brain* e *human brain*.



5.2 ANÁLISE DOS GENES CORRELACIONADOS COM DADOS EXPERIMENTAIS PUBLICADOS

5.2.1 DADOS DE TRANSCRIPTOMA

Os genes presentes nas listas mestres de humanos e camundongos foram analisados considerando dados publicados de expressão circadiana em larga escala. Para esta análise de evidência, atribuímos aos genes correlacionados nas “*listas mestres*” e identificados nestes estudos de transcriptoma circadiano, o valor de 0 para os genes que não apresentaram oscilação circadiana e 1 para os que oscilaram circadianamente. Aos genes que não apresentaram dados disponíveis foi atribuída a sigla ND (não disponível).

ZHANG e colaboradores (2014), caracterizaram a função do relógio circadiano no mecanismo fisiológico e comportamental de camundongos. Para isto, utilizaram as técnicas de *RNA-seq* e *Microarray* para quantificar o transcriptoma isoladamente ao longo do tempo de 12 órgãos/estruturas (cerebelo, corpo estriado, glândula adrenal, aorta, tronco cerebral, gordura marrom, coração, hipotálamo, rim, pulmão, fígado, músculo esquelético e gordura branca) de camundongo. Encontraram uma variação do transcriptoma a nível global e em determinados órgãos esta variação transcricional pode chegar até 43%. Estes dados de expressão estão disponíveis no banco de dados (Gene Expression Omnibus - GEO www.ncbi.nlm.nih.gov/geo, número de acesso: [GSE54652](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54652)), e foram gentilmente cedidos por John B. Hogenesch para nossas análises. A partir deste banco, explorou – se os dados dos genes que apresentaram ou não oscilação circadiana nas 12 estruturas analisadas, incluindo o cerebelo.

LI, J. Z. e colaboradores (2013), analisaram o padrão circadiano de expressão de genes em seis regiões do cérebro humano, sem histórico de doenças psiquiátricas ou neurológicas (grupo controle), no período menor que 11 horas após a morte, e no cérebro de pessoas com algum transtorno depressivo no período menor que 20 h após a morte. Eles ajustaram o registro da hora da morte, de cada sujeito, de acordo com o horário do nascer do sol do local em que ocorreu o óbito, realizando as análises circadianas em três ZTs (ZT0 = horário do nascer do sol, por volta das 06 horas da manhã, ZT 6 = meio dia e no ZT18= equivalente à meia noite). Utilizaram a técnica de *Microarray* para quantificar o transcriptoma de seis regiões cerebrais: cerebelo, hipocampo, córtex pré-frontal dorsolateral, córtex cingulado

anterior, amígdala e núcleo accumbens. Os dados demonstraram que mais de 100 genes comuns tiveram ritmicidade circadiana nas diferentes estruturas cerebrais estudadas. Em todo o cérebro foram identificados 3.573 genes com variação temporal. Em cerebelo, foram identificados 444 com ritmicidade de expressão. Os dados de expressão gênica estão disponibilizados no banco de dados (Gene Expression Omnibus GEO; www.ncbi.nlm.nih.gov/geo, número de acesso: [GSE45642](http://www.ncbi.nlm.nih.gov/geo/acc/GSE45642)), e também foram gentilmente cedidos por Jun Z. Li. Para as análises, utilizamos os dados do grupo controle.

MARET e colaboradores (2007), com o objetivo demonstrar uma relação entre fatores genéticos e o ritmo do sono, realizaram experimentos com privação de sono e análises do perfil circadiano do transcriptoma no cérebro e fígado de camundongos, através das técnicas de microarray e RT-PCR. Neste estudo, trezentos e noventa e um transcritos permanecem rítmicos em camundongos privados de sono (grupo experimental) e 2.032 transcritos possuem expressão circadiana no grupo controle. Estes dados estão disponíveis *on line* no material suplementar do artigo, tabela 03 (<http://www.pnas.org/content/104/50/20090?tab=ds#ST>). Utilizamos os resultados de expressão obtida no grupo controle, para avaliar se os genes correlacionados apresentaram variação circadiana no tecido investigado.

HUGHES e colaboradores (2009), realizaram uma comparação do perfil circadiano de transcriptoma do fígado de camundongos, e de duas linhagens celulares (NIH3T3 e U2OS), utilizando análises de *Microarray* e RT-PCR. Dentre os diversos resultados desta pesquisa, foi demonstrado que 3.667 transcritos no fígado e menos de 12 nas linhagens celulares possuem ritmo de expressão em torno de 24 horas. Estes dados estão disponíveis *on line* no material suplementar do artigo, tabela 01 (<http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000442#s5>).

5.2.2 DADOS DE PROTEOMA

Com base em dados de proteoma, avaliamos se os genes correlacionados possuem ou não evidência de oscilação em suas respectivas proteínas. Para os genes que apresentam oscilação a nível proteico foram-lhes atribuídos os valores de 1, para os que não apresentam foram-lhes atribuídos os valores de 0 e aqueles que não possuem dados disponíveis, foram identificados pela sigla ND (não disponível).

ROBLES e colaboradores (2014), quantificaram o proteoma circadiano em amostras de fígados de camundongos em 16 horários ao longo de dois dias. Os autores relataram que das

quase 3000 proteínas analisadas, 186 apresentaram oscilação circadiana. Os dados das proteínas que apresentaram e não apresentaram oscilação circadiana estão disponíveis no material suplementar do artigo, tabela 01 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3879213/>).

MAUVOISIN e colaboradores (2014) desenvolveram um estudo similar, no qual analisaram o proteoma de fígado de camundongos em 16 momentos ao longo de dois dias. Encontraram que 195 proteínas possuem ritmicidade circadiana. Esses dados estão disponíveis no material suplementar do artigo, tabela 01 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3890886/>).

CHIANG e colaboradores (2014), analisaram o proteoma circadiano em *NSQ* de camundongos, em seis horários durante um dia. Quantificaram 2112 proteínas, das quais 421 apresentaram oscilação circadiana. Estes resultados estão disponíveis no material suplementar do artigo, tabela 01 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4199512/>).

5.2.3 DADOS DE *CHIP-SEQ*

Confrontamos os genes correlacionados obtidos em nossa pesquisa com os estudos de *ChiP-seq*, considerando a evidência experimental de sítio de ligação para algum fator de transcrição circadiano, dando-lhes o valor de 1 para os genes que foram identificados no *ChiP-seq* para algum dos fatores de transcrição e 0 para os genes que não apresentaram evidência de sítios de ligação. Cada evidência de *ChiP-seq* foi considerada independentemente, ou seja, o valor de 1 atribuído para cada fator de transcrição.

KOIKE e colaboradores (2013) analisaram a arquitetura transcricional da alça regulatória numa escala genômica do cistroma circadiano em fígado de camundongo. Para isto, realizaram análises de *ChiP-seq* para sete fatores de transcrição circadianos, BMAL1, CLOCK, NPAS, CRY1, CRY2, PER1 e PER2 e identificaram os genes que possuem sítios de ligação para estes fatores. Os resultados para cada fator de transcrição estão presentes no material suplementar do artigo, tabela 02 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3694775/>).

FANG e colaboradores (2014), para avaliar a funcionalidade do gene *Nr1d1* no cistroma circadiano, realizaram análises de *ChiP-seq* em algumas fases do dia, em fígado de camundongo *Knockout* para este gene. Eles encontraram que diversos genes possuem sítios de

ligação para o fator de transcrição NR1D1, tendo seus transcritos inibidos ou ativados após a deleção do mesmo. Estes dados estão disponíveis no material suplementar do artigo tabela 04 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4243056/>).

5.2.3 DADOS DE INTERAÇÃO PROTEÍNA-PROTEÍNA

Foi utilizado o portal *Protein Network Analysis* (PINA-<http://cbg.garvan.unsw.edu.au/pina/>) para coletar dados de interação de proteína nas espécies *M. musculus* e *H. sapiens* com as 19 proteínas do relógio molecular analisadas (BMAL1, BMAL2, CLOCK, CRY1, CRY2, CSNK1A, CSNK1D, CSNK1E, FBLX3, FBLX21, NPAS2, NR1D1, NR1D2, PER1, PER2, PER3, RORA, RORB, E RORC). Assim, foram identificados nas listas de correlação os genes cujas proteínas codificadas apresentaram ou não interação com as proteínas do relógio. Foi atribuído os valores 1 e 0 para presença e ausência de interação, respectivamente.

5.2.4 DADOS DE RNA DE INTERFERÊNCIA (RNAI)

ZHANG e colaboradores (2009), realizaram um estudo em escala genômica com RNAi em células humanas. Os autores mediram a expressão de gene repórter de luciferase (BMAL1 e *PER2 - dLuc*) a cada duas horas durante 4 dias. Detectaram que aproximadamente 250 genes, demonstraram mudanças nos níveis de expressão, em amplitude ou período do ritmo bioluminescente do gene reporter. Estes dados estão disponíveis *on line* no banco de dados (*PubChem Bioassay do National Center for Biotechnology Information*). Para os genes que foram associados com mudanças de amplitude e de período do ritmo, atribuímos o valor 1, para os que não apresentaram associação com mudanças no fenótipo circadiano, de acordo com o estudo de ZHANG et al. (2009) atribuímos o valor 0.

5.3 ANÁLISE DE REDES DE INTERAÇÃO FUNCIONAL PROTEÍNA-PROTEÍNA.

Através do portal *on line* STRING (<http://string-db.org>), que fornece dados de associações entre proteínas, incluindo associações diretas (físicas) e indiretas (funcionais) estabelecendo redes de conexões (SZKLARCZYK, Damian et al.2014). Através deste portal, realizamos análises de interações funcionais das 19 proteínas circadianas estudadas, com 2% das proteínas correspondentes aos genes com os maiores valores de correlações positivas e

negativas em cerebelo de *M. musculus* e para 2% das proteínas, correspondentes aos genes com os menores valores de correlações positivas e negativas.

5.4 ANÁLISES ESTATÍSTICAS

Para análise estatística dos dados, os genes identificados nas listas mestres foram inicialmente segregados em grupos que apresentam ou não oscilação circadiana de acordo com os estudos de ZANG et al. (2014) e LI et al. (2013). Realizamos então, uma análise da frequência relativa em intervalos de faixas de 0,1 dos valores de correlações encontrados apenas entre os 19 genes circadianos e do conjunto total de correlações encontrado para os 19 genes circadianos avaliados em cerebelo de *M. musculus* e *H. sapiens*. A frequência relativa nestes grupos foi gerada com base no quantitativo de correlações encontradas por faixas em relação ao quantitativo total de correlações encontradas em cada grupo. Para isto, utilizamos o Software IBM SPSS Statistics versão 20 e os gráficos foram gerados através do Excel, 2010.

Posteriormente, segregamos os dados em grupos de acordo com a presença ou ausência de evidências de associação com o ritmo circadiano. Para análise dos grupos que apresentaram e não apresentaram variação circadiana em cerebelo (dados de transcriptoma de cerebelo obtidos de ZHANG et al.,2014 e Li, J. Z. et al.,2013) , avaliamos a frequência relativa da soma desses dois grupos por faixas de correlações, com intervalos de 0,1. Foi utilizado o teste de *Mann Whitney para* comparar as médias dos valores de correlação de *Pearson* em módulo ($|r|$), ou seja, considerando o valor numérico absoluto das correlações desprezando seus sinais, dos grupos de genes que apresentaram e não apresentaram oscilação circadiana. Para análise de outras evidências (transcriptoma circadiano em outros tecidos, proteoma, ChiP-seq, interação proteína-proteína e RNAi), selecionamos 2% dos genes com valores mais extremos de r positivos e negativos e avaliamos a frequência de genes que apresentam ou não evidências em cada grupo e cada categoria de experimentos, utilizando o teste do Qui-quadrado. Estas análises foram realizadas por meio do *Software GraphPad Prims* versão 6. O nível de significância adotado foi de $p < 0,05$.

6 RESULTADOS

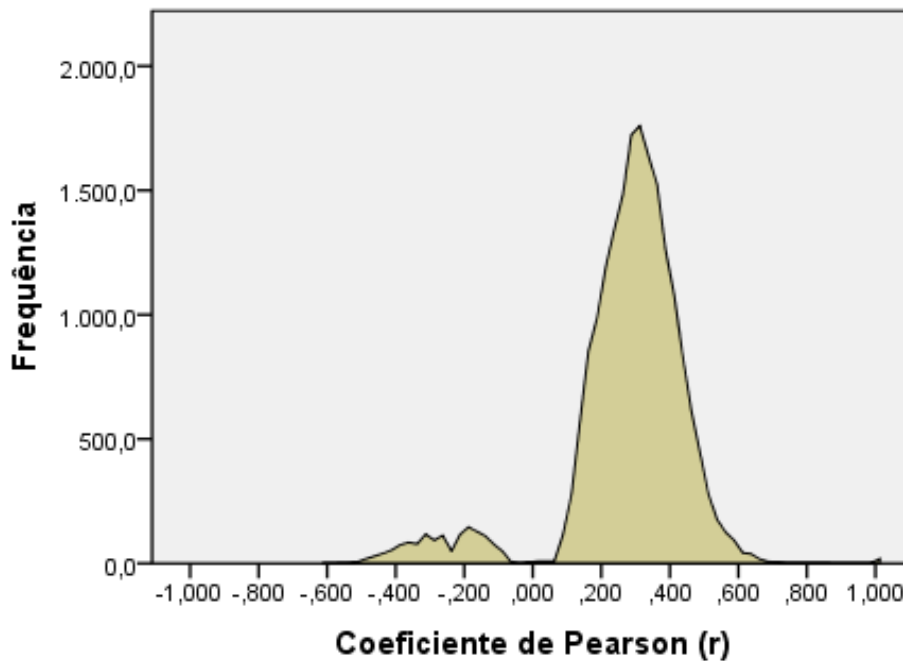
6.1 IDENTIFICAÇÃO DE GENES CORRELACIONADOS A GENES CIRCADIANOS NO CEREBELO DE HUMANOS E DE CAMUNDONGOS A PARTIR DO *ALLEN BRAIN ATLAS*

Obtivemos 19.907 genes no cerebelo de *M. musculus* e 29.176 genes no cerebelo de *H. sapiens*, cujos padrões de expressão foram correlacionados com um ou mais dos 19 genes circadianos investigados. Considerando estes dois conjuntos de dados, 14.242 genes são comuns entre as duas espécies. A quantidade de correlações ou valores de r , encontradas nesses três grupos corresponde à quantidade de genes em cada um, ou seja, 19.907 correlações no cerebelo de *M. musculus*, 29.176 no cerebelo de *H. sapiens* e 14.242 entre as duas espécies.

Os valores de correlações são variados e não indicam, a priori, altas correlações para todos os genes, tendo em vista que estão distribuídos entre os valores de -1 a +1. No entanto, o gene *Pkia* (cAMP-dependent protein kinase inhibitor alpha), apresentou valor de r acima de 0,9 nos nossos resultados para *Mus musculus*. Porém, devido a informação dada pelo Portal *Allen Brain Atlas*, de que houve problemas na sonda para este gene utilizada na técnica de *HIS*, o mesmo foi excluído dos dados.

As figuras 10 e 11 apresentam a distribuição nos conjuntos de dados para camundongos e humanos dos maiores valores de correlação que um dado gene apresentou com um gene relógio. Os dados de correlação positiva para *M. musculus* variam de 0,0 a 0,729, com média de 0,312 e mediana de 0,309, sendo que a maior frequência de genes possui valores de r 0,302 (Figura 10). Os dados de correlação negativa possuem valores de r entre -0,026 a -0,633, com média de -0,357 e mediana de -0,252, sendo que a maior frequência de genes possui valores de r de -0,169.

Figura 10. Distribuição dos 19.907 genes correlacionados com os genes circadianos no cerebelo de *M. musculus*, de acordo com seus valores de r que variam de -0,633 a 0,729.

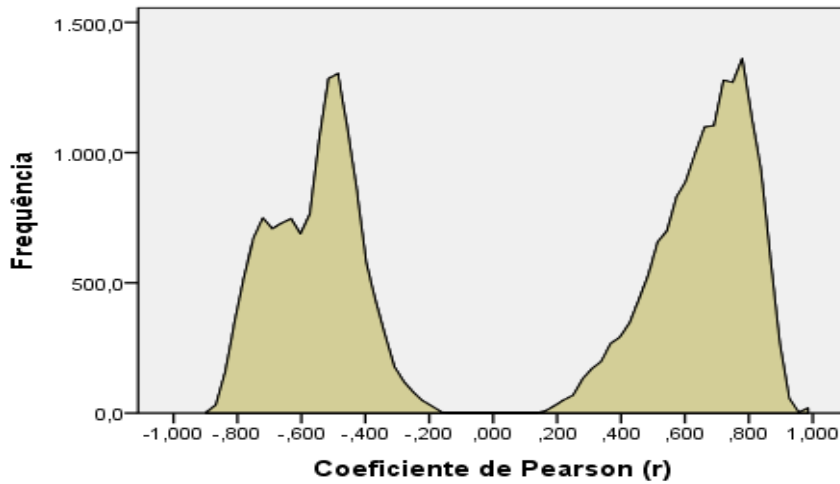


Fonte: Autor

Para *H. sapiens*, os valores dos dados de correlação positiva possuem valores de r entre 0,13 a 0,948, com média de 0,656 e mediana de 0,681, sendo que a maior frequência de genes possui valores de r de 0,781. Os dados de correlação negativa possuem valores de r entre -0,147 a -0,899, com média de -0,753 e mediana de -0,54, sendo que a maior frequência de genes possui valores de -0,506 (figura 11).

Identificamos para os genes homólogos um total de 13.258 genes com valores de r positivos e 984 genes com valores de r negativos, para a espécie *M. musculus* e 8.129 genes com valores de r positivos e 6.113 com valores negativos para espécie *H. sapiens*

Figura 11. Distribuição dos 29.176 genes correlacionados com os genes circadianos no cerebelo de *H. sapiens*, de acordo com seus valores de r que variam de - 0,899 a 0,948.



Fonte: Autor

Foi obtida, desta forma, uma lista unificada de genes correlacionados a genes circadianos, cada um com o maior valor de correlação identificado. As tabelas 02 e 03 a seguir representam uma parcela desses genes ranqueados de acordo com os maiores valores de correlação, independentemente de serem positivos ou negativos. Observa-se que para cada gene correlacionado nas espécies *M. musculus* e *H. sapiens*, os valores de r são variados nos diferentes genes relógio. Neste caso, o maior valor de r (presente na segunda coluna das tabelas 01 e 02) para cada gene correlacionado, seja positivo ou negativo, foi preservado.

Tabela 02. Valores de correlação de Pearson (r) de 10 genes com expressão correlacionada com os 19 genes circadianos, em cerebelo de camundongos. Os valores acima de 0,6 estão destacados em negrito.

Genes	r/p	<i>Bmal1</i>	<i>Bmal2</i>	<i>Clock</i>	<i>Cry1</i>	<i>Cry2</i>	<i>Csnk1a</i>	<i>Csnk1d</i>	<i>Csnk1e</i>	<i>Fbxl21</i>	<i>Fbxl3</i>	<i>Npas2</i>	<i>Nr1d1</i>	<i>Nr1d2</i>	<i>Per1</i>	<i>Per2</i>	<i>Per3</i>	<i>Rora</i>	<i>Rorb</i>	<i>Rorc</i>
<i>Cartpt</i>	0,729	-0,156	0,049	0,096	-0,022	0,119	0,216	0,046	0,062	0,032	-0,059	0,079	-0,177	-0,157	0,139	0,148	0,145	-0,14	0,729	0,036
<i>Gpr149</i>	0,725	0,197	0,268	0,191	-0,004	0,244	0,054	-0,04	0,089	0,134	0,182	0,725	0,249	0,168	0,38	0,183	-0,092	0,24	-0,073	-0,048
<i>Inpp4a</i>	0,721	0,721	-0,192	0,131	-0,011	0,222	0,174	-0,152	-0,209	0,057	0,452	0,127	0,399	0,202	0,26	0,19	-0,149	0,636	-0,098	-0,034
<i>Gal</i>	0,713	0,202	0,141	0,062	-0,026	-0,055	0,225	0,155	0,197	0,069	-0,009	0,18	-0,147	0,084	-0,037	0,092	0,169	-0,168	0,713	0,088
<i>Slc5a7</i>	0,711	-0,111	0,063	-0,073	-0,022	-0,048	0,207	0,077	0,095	-0,072	-0,017	0,076	-0,117	-0,112	0,09	0,108	0,184	-0,159	0,711	0,04
<i>Shisa6</i>	0,702	0,642	-0,202	0,191	-0,037	0,261	0,255	-0,177	-0,375	-0,02	0,537	0,06	0,371	0,221	0,352	0,197	-0,07	0,702	-0,085	0,027
<i>Pggt1b</i>	0,698	0,335	-0,047	-0,062	-0,003	0,083	-0,057	0,046	0,078	0,698	0,092	0,115	0,252	0,156	0,144	0,133	-0,063	0,084	-0,143	-0,069
<i>Gng13</i>	0,684	0,607	-0,158	0,219	-0,026	0,429	0,265	-0,202	-0,398	0,024	0,478	0,147	0,347	0,2	0,527	0,229	0,116	0,684	-0,16	0,077
<i>Grid2</i>	0,678	0,57	-0,15	0,237	0,042	0,296	0,241	-0,161	-0,372	0,073	0,498	0,167	0,385	0,168	0,294	0,2	-0,106	0,678	-0,126	0,062
<i>Calb1</i>	0,676	0,612	-0,247	0,254	0,047	0,389	0,324	-0,201	-0,388	-0,072	0,552	0,238	0,415	0,263	0,436	0,193	-0,117	0,676	-0,144	0,074

Fonte: Autor

*r/p= valor de correlação de Pearson preservado.

Tabela 03. Valores de correlação de Pearson (r) de 10 genes com expressão correlacionada com os 19 genes circadianos, em cerebelo de humanos. Os valores acima de 0,6 estão destacados em negrito

Genes	r/p	<i>Bmal1</i>	<i>Bmal2</i>	<i>Clock</i>	<i>Cry1</i>	<i>Cry2</i>	<i>Csnk1a</i>	<i>Csnk1d</i>	<i>Csnk1e</i>	<i>Fbxl21</i>	<i>Fbxl3</i>	<i>Npas2</i>	<i>Nr1d1</i>	<i>Nr1d2</i>	<i>Per1</i>	<i>Per2</i>	<i>Per3</i>	<i>Rora</i>	<i>Rorb</i>	<i>Rorc</i>
<i>Clorf21</i>	0,948	0,665	0,679	0,687	0,867	0,948	0,791	0,541	0,453	0,276	-0,438	0,197	0,456	0,663	-0,47	0,847	0,768	0,573	-0,241	0,61
<i>Nfia</i>	0,945	0,676	0,68	0,694	0,81	0,945	0,859	0,449	-0,117	0,386	0,179	0,596	0,682	0,396	0,831	0,771	0,625	0,147	0,309	-0,477
<i>Hnrnpa3</i>	0,942	0,673	0,784	0,69	0,85	0,942	0,884	-0,547	0,502	0,31	0,741	-0,211	0,466	0,65	0,411	0,829	0,754	0,581	-0,306	-0,398
<i>Ccdc85c</i>	0,94	0,753	0,751	0,74	0,846	0,94	0,808	-0,582	0,517	0,157	0,272	0,188	0,598	0,552	0,528	0,784	0,728	0,595	-0,381	0,406
<i>Fzd3</i>	0,934	0,656	0,709	0,668	0,864	0,934	0,775	-0,56	0,359	0,125	0,398	0,189	0,447	0,622	0,362	0,805	0,715	0,622	-0,206	-0,412
<i>Zic2</i>	0,933	0,718	0,793	0,647	0,933	0,826	0,809	-0,378	0,321	0,241	0,12	-0,179	0,195	0,579	0,139	0,687	0,563	0,575	-0,174	-0,351
<i>Fut9</i>	0,927	0,833	0,596	0,769	0,825	0,927	0,644	-0,704	0,543	0,165	0,107	-0,304	0,668	0,744	0,57	0,855	0,88	0,684	-0,246	-0,365
<i>Pdk1</i>	0,927	0,785	0,631	0,668	0,694	0,819	0,582	-0,715	0,593	0,346	0,604	-0,28	0,69	0,767	0,658	0,819	0,927	0,563	-0,195	-0,439
<i>Tubgcp6</i>	0,926	0,735	0,352	0,612	0,614	0,778	-0,512	-0,664	0,926	-0,197	-0,382	0,302	0,772	0,383	0,756	0,729	0,749	0,351	-0,426	0,631
<i>C14orf132</i>	0,925	0,811	0,631	0,777	0,793	0,925	0,672	-0,636	0,568	-0,1	-0,085	0,259	0,699	0,64	0,611	0,891	0,846	0,608	-0,322	0,485

Fonte: Autor

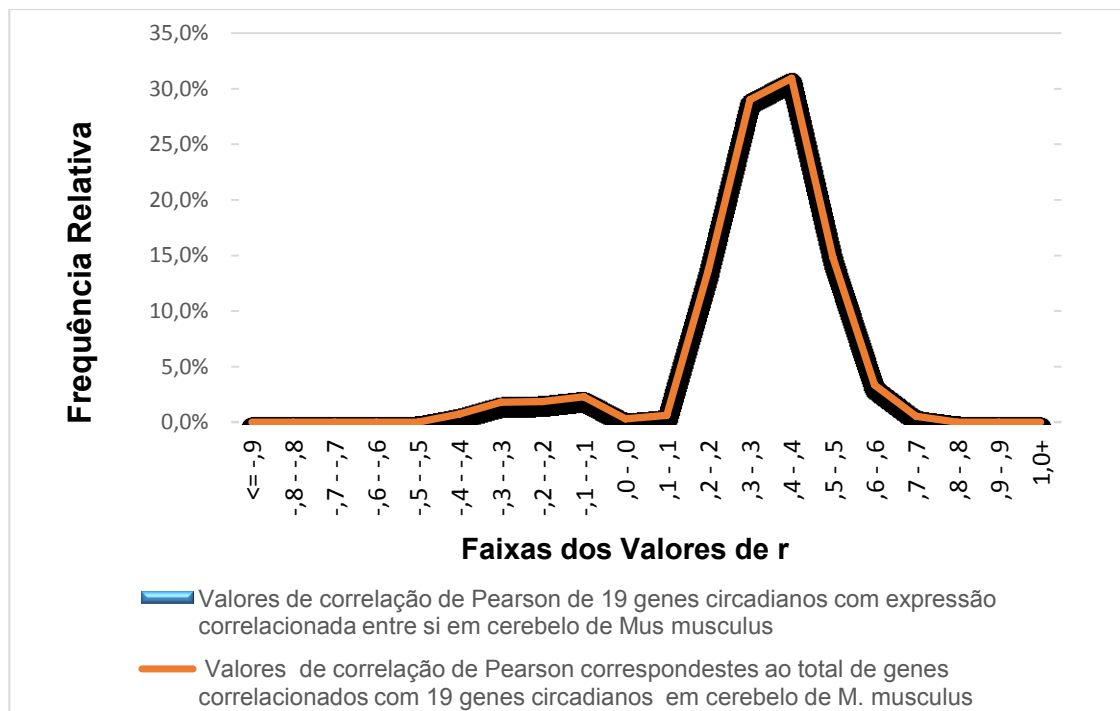
*r/p= valor de correlação de Pearson preservado.

6.2 ANÁLISE DA CORRELAÇÃO ENTRE OS 19 GENES CIRCADIANOS EM CEREBELO DE *MUS MUSCULUS* E *H. SAPIENS*.

Foram obtidos 171 valores únicos de correlações entre os 19 genes circadianos em cerebelo da espécie *M. musculus*, dos quais 139 valores são positivos dentro de uma faixa de 0 a 0,628, média de 0,184 e mediana de 0,162, e 32 valores negativos que possuem valores entre -0,007 a -0,329, com média de -0,114 e mediana de -0,083. O perfil de distribuição destes 171 valores de correlação e os 19.907 genes, correspondentes ao total de genes correlacionados encontrados em cerebelo de *M. musculus*, está apresentada na (Figura12).

A maior frequência de correlações do grupo de genes circadianos encontra-se na faixa entre r 0,2 e 0,299. Para o grupo geral de correlações, a maior frequência de genes encontra-se na faixa de 0,4 a 0,499 de valores de r (Figura 12).

Figura 12. Distribuição por Frequência relativa (em relação ao total de cada grupo) dos valores de correlação de *Pearson* de 19 genes circadianos com expressão correlacionada entre si e do total de genes correlacionados com 19 genes circadianos obtidos em cerebelo de *M. musculus*, por faixas de 0,1.



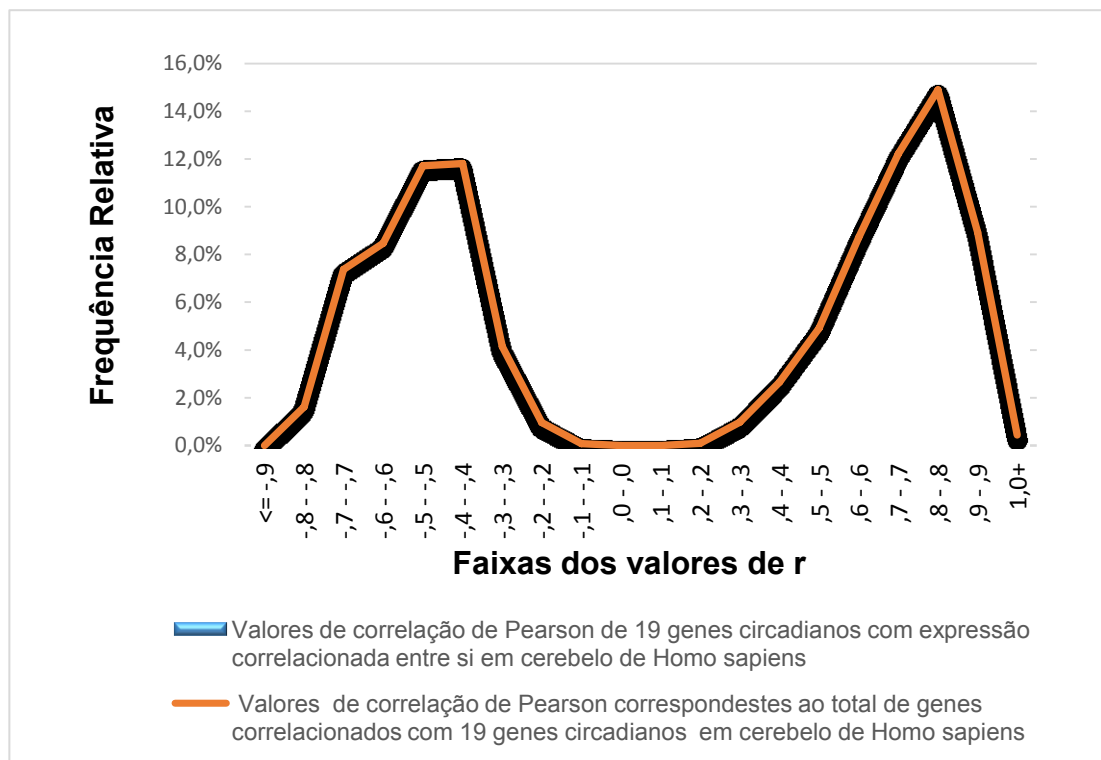
Identificamos também, os 171 valores únicos de correlações obtidos entre os 19 genes circadianos em cerebelo da espécie *H. Sapiens*, dos quais 115 são positivos dentro de uma faixa 0,096 a 0,859, com média de 0,498 e mediana de 0,522, e 56 valores negativos que possuem valores entre -0,076 a -0,809, com média de -0,347 e mediana de -0,083. O perfil de

distribuição destes valores em termos de frequência relativa ao total de cada grupo está representado na figura 13.

A maior frequência de correlações do grupo de genes circadianos encontra-se na faixa entre r 0,7 e 0,799. Para o grupo geral de correlações, a maior frequência de genes encontra-se na faixa de 0,8 a 0,899 de valores de r (Figura 13).

O grupo de correlações geral, com todos os genes identificados correlacionados a genes circadianos, apresentou maiores frequências em faixas de valores maiores, sejam negativos ou positivos, comparados ao grupo de correlações entre os genes circadianos (Figura 13). Além disto, o grupo de correlações entre genes circadianos apresentou uma distribuição mais homogênea dos dados.

Figura 13. Distribuição por Frequência relativa (em relação ao total de cada grupo) dos valores de correlação de *Pearson* de 19 genes circadianos com expressão correlacionada entre si e do total de genes correlacionados com 19 genes circadianos obtidos em cerebelo de *H. sapiens*, por faixas de 0,1.



As tabelas 04 e 05 a seguir representam os valores das correlações de *Pearson* de 19 genes circadianos com expressão espacial correlacionada entre si, em cerebelo de camundongos e humanos. Observou-se que nas correlações entre esses genes, tanto na espécie *M. musculus*, quanto em *H. sapiens*, os valores de *r* são variados de acordo com o gene circadiano correlacionado. Neste caso também, o maior valor de *r* para cada gene circadiano correlacionado, seja positivo ou negativo, foi preservado.

Além disso, identificamos 14 valores com correlações médias e altas entres os 19 genes circadianos estudados, que variam de 0,407 a 0,628, em cerebelo de *M. musculus*, sendo os três maiores valores de 0,628; 0,603 e 0,531 encontrados respectivamente, entre os genes *Per2* e *Cry2*; *Bmal1* e *Rora* e *Rora* e *Fbxl3*. Em *H. sapiens*, foram identificados 28 valores com correlações altas, que variam de -0,706 a -0,809 e de 0,702 a 0,859 em cerebelo de *H. sapiens*. Seis correlações apresentaram valores acima de 0,8: *Per2* e *Per3* (0,859), *Per2* e *Cry2* (0,855), *Cry1* e *Cry2* (0,838), *Per1* e *Nr1d1* (0,822), *Csnk1a* e *Bmal1* (0,808) e *Per1* e *Csnk1d* (-0,809).

Tabela 04. Valores de correlação de Pearson (r) de 19 genes circadianos com expressão correlacionada entre si, em cerebelo de camundongos. Os valores acima de 0,6 estão destacados em negrito.

	<i>Bmal1</i>	<i>Bmal2</i>	<i>Clock</i>	<i>Cry1</i>	<i>Cry2</i>	<i>Csnk1a</i>	<i>Csnk1d</i>	<i>Csnk1e</i>	<i>Fbxl3</i>	<i>Fbxl21</i>	<i>Npas2</i>	<i>Nr1d1</i>	<i>Nr1d2</i>	<i>Per1</i>	<i>Per2</i>	<i>Per3</i>	<i>Rora</i>	<i>Rorb</i>	<i>Rorc</i>
<i>Bmal1</i>	1	0,095	0,151	-0,007	0,436	0,206	0,222	0,168	0,454	0,169	0,162	0,407	0,205	0,321	0,33	0,056	0,628	-0,102	0,072
<i>Bmal2</i>	0,095	1	-0,051	0,01	0,139	-0,083	0,223	0,161	-0,133	-0,06	0,278	-0,107	-0,136	0,113	0,263	0,034	-0,276	0,112	0,007
<i>Clock</i>	0,151	-0,051	1	0,01	0,247	0,108	0,129	-0,113	0,178	-0,07	0,064	0,202	0,233	0,226	0,193	0,055	0,199	0,057	0,111
<i>Cry1</i>	-0,007	0,01	0,01	1	-0,044	0,022	0,026	-0,033	0,04	0,02	-0,015	0,049	0,038	0,018	0,085	0,016	-0,068	0,033	0
<i>Cry2</i>	0,436	0,139	0,247	-0,044	1	0,288	-0,078	-0,199	0,396	0,017	0,186	0,453	0,193	0,496	0,603	0,117	0,511	0,217	-0,037
<i>Csnk1a</i>	0,206	-0,083	0,108	0,022	0,288	1	0,169	0,147	0,392	0,012	0,088	0,246	0,234	0,317	0,22	0,252	0,331	0,226	0,161
<i>Csnk1d</i>	0,222	0,223	0,129	0,026	-0,078	0,169	1	0,176	-0,083	-0,018	0,167	0,102	0,23	0,136	0,167	0,19	-0,182	0,15	0,101
<i>Csnk1e</i>	0,168	0,161	-0,113	-0,033	-0,199	0,147	0,176	1	-0,158	0,089	0,282	-0,299	0,185	0,128	0,107	0,088	-0,329	0,151	0,044
<i>Fbxl3</i>	0,454	-0,133	0,178	0,04	0,396	0,392	-0,083	-0,158	1	0,124	0,106	0,321	0,182	0,416	0,228	0,077	0,531	0,131	0,081
<i>Fbxl21</i>	0,169	-0,06	-0,07	0,02	0,017	0,012	-0,018	0,089	0,124	1	0,075	0,206	0,123	0,153	0,01	0,014	-0,217	-0,077	-0,08
<i>Npas2</i>	0,162	0,278	0,064	-0,015	0,186	0,088	0,167	0,282	0,106	0,075	1	0,219	-0,189	0,218	0,408	0,194	0,181	0,116	-0,034
<i>Nr1d1</i>	0,407	-0,107	0,202	0,049	0,453	0,246	0,102	-0,299	0,321	0,206	0,219	1	0,214	0,421	0,51	-0,082	0,415	0,171	0,083
<i>Nr1d2</i>	0,205	-0,136	0,233	0,038	0,193	0,234	0,23	0,185	0,182	0,123	-0,189	0,214	1	0,15	-0,143	0,113	0,224	0,141	0,066
<i>Per1</i>	0,321	0,113	0,226	0,018	0,496	0,317	0,136	0,128	0,416	0,153	0,218	0,421	0,15	1	0,294	0,081	0,437	0,17	0,098
<i>Per2</i>	0,33	0,263	0,193	0,085	0,603	0,22	0,167	0,107	0,228	0,01	0,408	0,51	-0,143	0,294	1	0,064	0,25	0,278	0,119
<i>Per3</i>	0,056	0,034	0,055	0,016	0,117	0,252	0,19	0,088	0,077	0,014	0,194	-0,082	0,113	0,081	0,064	1	-0,162	0,17	0,143
<i>Rora</i>	0,628	-0,276	0,199	-0,068	0,511	0,331	-0,182	-0,329	0,531	-0,217	0,181	0,415	0,224	0,437	0,25	-0,162	1	0,136	0,121
<i>Rorb</i>	-0,102	0,112	0,057	0,033	0,217	0,226	0,15	0,151	0,131	-0,077	0,116	0,171	0,141	0,17	0,278	0,17	0,136	1	0,102
<i>Rorc</i>	0,072	0,007	0,111	0	-0,037	0,161	0,101	0,044	0,081	-0,08	-0,034	0,083	0,066	0,098	0,119	0,143	0,121	0,102	1

Fonte: Autor

Tabela 05. Valores de correlação de Pearson (r) de 19 genes circadianos com expressão correlacionada entre si, em cerebelo de humanos. Os valores acima de 0,6 estão destacados em negrito.

	<i>Bmal1</i>	<i>Bmal2</i>	<i>Clock</i>	<i>Cry1</i>	<i>Cry2</i>	<i>Csnk1a</i>	<i>Csnk1d</i>	<i>Csnk1e</i>	<i>Fbxl3</i>	<i>Fbxl21</i>	<i>Npas2</i>	<i>Nr1d1</i>	<i>Nr1d2</i>	<i>Per1</i>	<i>Per2</i>	<i>Per3</i>	<i>Rora</i>	<i>Rorb</i>	<i>Rorc</i>
<i>Bmal1</i>	1	0,756	0,686	0,766	0,774	0,59	-0,732	0,553	0,334	0,161	-0,232	0,732	0,569	0,653	0,717	0,8	0,651	-0,231	0,399
<i>Bmal2</i>	0,555	1	0,702	0,756	0,621	0,808	0,267	0,342	0,511	0,268	0,253	-0,319	0,488	-0,311	0,559	0,442	0,514	-0,132	-0,376
<i>Clock</i>	0,686	0,702	1	0,658	0,718	0,692	-0,396	0,381	0,522	0,227	0,131	0,548	0,551	0,343	0,618	0,625	0,583	-0,249	-0,327
<i>Cry1</i>	0,766	0,756	0,658	1	0,838	0,761	-0,428	0,632	0,168	0,287	-0,175	0,304	0,672	0,407	0,727	0,619	0,604	-0,226	-0,383
<i>Cry2</i>	0,774	0,621	0,718	0,838	1	0,729	-0,706	0,545	-0,193	0,127	0,325	0,688	0,65	0,764	0,855	0,794	0,636	-0,251	0,58
<i>Csnk1a</i>	0,59	0,808	0,692	0,761	0,729	1	0,474	-0,459	0,564	0,239	0,242	-0,501	0,498	-0,459	0,637	0,488	0,506	0,327	-0,432
<i>Csnk1d</i>	-0,732	0,267	-0,396	-0,428	-0,706	0,474	1	-0,584	0,384	0,115	0,301	-0,744	-0,517	-0,809	-0,652	-0,781	-0,254	0,129	-0,528
<i>Csnk1e</i>	0,553	0,342	0,381	0,632	0,545	-0,459	-0,584	1	-0,388	-0,353	0,251	0,504	0,575	0,718	0,658	0,636	0,265	-0,442	0,457
<i>Fbxl3</i>	0,334	0,511	0,522	0,168	-0,193	0,564	0,384	-0,388	1	0,096	-0,129	-0,208	0,434	-0,412	-0,148	0,343	0,265	0,52	-0,299
<i>Fbxl21</i>	0,161	0,268	0,227	0,287	0,127	0,239	0,115	-0,353	0,096	1	0,162	-0,331	0,226	-0,151	0,134	0,152	0,25	-0,076	-0,286
<i>Npas2</i>	-0,232	0,253	0,131	-0,175	0,325	0,242	0,301	0,251	-0,129	0,162	1	-0,222	-0,362	-0,313	-0,315	-0,322	0,107	-0,138	0,137
<i>Nr1d1</i>	0,732	-0,319	0,548	0,304	0,688	-0,501	-0,744	0,504	-0,208	-0,331	-0,222	1	0,43	0,822	0,612	0,781	0,315	-0,122	0,639
<i>Nr1d2</i>	0,569	0,488	0,551	0,672	0,65	0,498	-0,517	0,575	0,434	0,226	-0,362	0,43	1	0,516	0,681	0,74	0,514	-0,092	-0,366
<i>Per1</i>	0,653	-0,311	0,343	0,407	0,764	-0,459	-0,809	0,718	-0,412	-0,151	-0,313	0,822	0,516	1	0,709	0,759	0,104	-0,289	0,66
<i>Per2</i>	0,717	0,559	0,618	0,727	0,855	0,637	-0,652	0,658	-0,148	0,134	-0,315	0,612	0,681	0,709	1	0,859	0,496	-0,328	-0,507
<i>Per3</i>	0,8	0,442	0,625	0,619	0,794	0,488	-0,781	0,636	0,343	0,152	-0,322	0,781	0,74	0,759	0,859	1	0,459	-0,189	0,472
<i>Rora</i>	0,651	0,514	0,583	0,604	0,636	0,506	-0,254	0,265	0,265	0,25	0,107	0,315	0,514	0,104	0,496	0,459	1	0,139	-0,089
<i>Rorb</i>	-0,231	-0,132	-0,249	-0,226	-0,251	0,327	0,129	-0,442	0,52	-0,076	-0,138	-0,122	-0,092	-0,289	-0,328	-0,189	0,139	1	-0,187
<i>Rorc</i>	0,399	-0,376	-0,327	-0,383	0,58	-0,432	-0,528	0,457	-0,299	-0,286	0,137	0,639	-0,366	0,66	-0,507	0,472	-0,089	-0,187	1

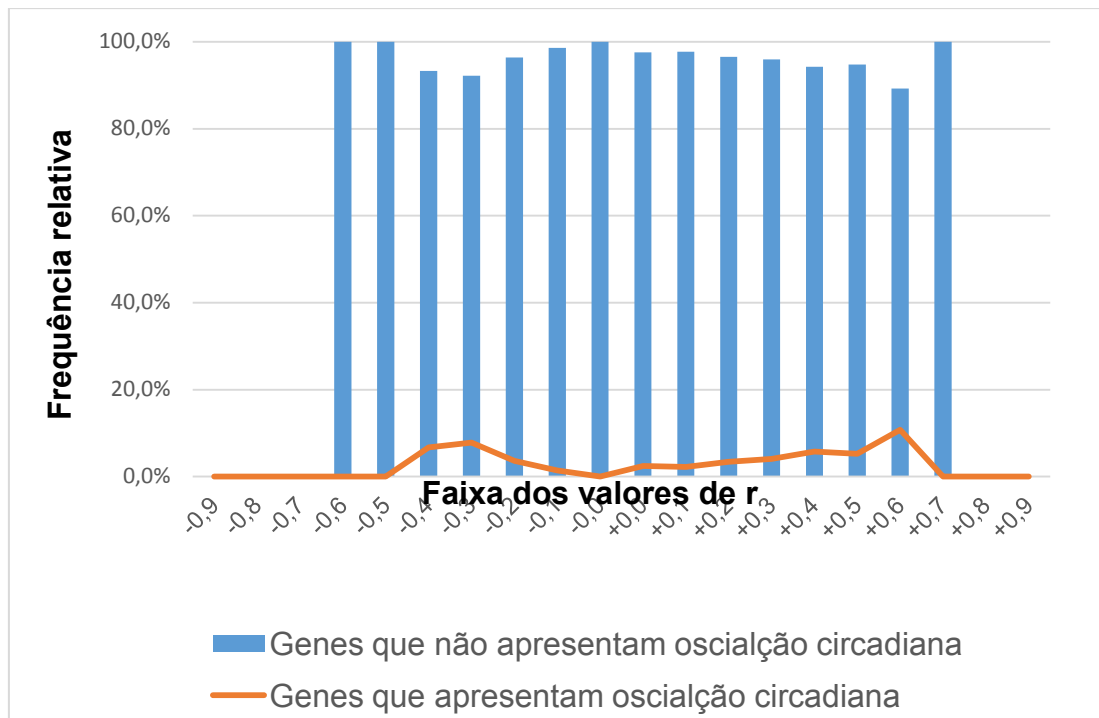
Fonte: Autor

6.3 Validação dos dados com base em estudos de transcriptoma circadiano de cerebelo

Os resultados do quantitativo total de correlações e genes obtidos, ou seja, 19.907 no cerebelo de *M. musculus* e 29.176 genes no cerebelo de *H. sapiens*, foram confrontados com dados de estudos cronobiológicos de transcriptoma em larga escala realizados em amostras de cerebelo murino (ZHANG et al., 2014) e humano (LI et al 2013), para determinar, dentre os genes correlacionados, aqueles que haviam sido previamente identificados com uma variação circadiana de expressão. Desta forma, identificamos que no grupo composto por 19.907 correlacionados em cerebelo de camundongos, 16.799 genes (84,3 %) estão presentes nos dados de ZHANG et al., 2014, dos quais 476 (4,03%) apresentaram oscilação circadiana (Figura 14). A proporção de genes que apresentou variação circadiana em cerebelo, considerando apenas os dados de ZHANG et al., 2014 foi de 3,67% (726 de um total de 19788 genes analisados). Identificamos também, que no grupo composto por 29.176 correlacionados em cerebelo de humanos, 6.003 genes (20,5 %) estão presentes nos dados de LI et al., 2013, dos quais 414 (6,8%) apresentaram oscilação circadiana (Figura 15). Considerando apenas os dados de LI et al., 2013, 3,72% dos genes avaliados apresentaram variação circadiana no cerebelo(444 de 11.912 genes avaliados) .

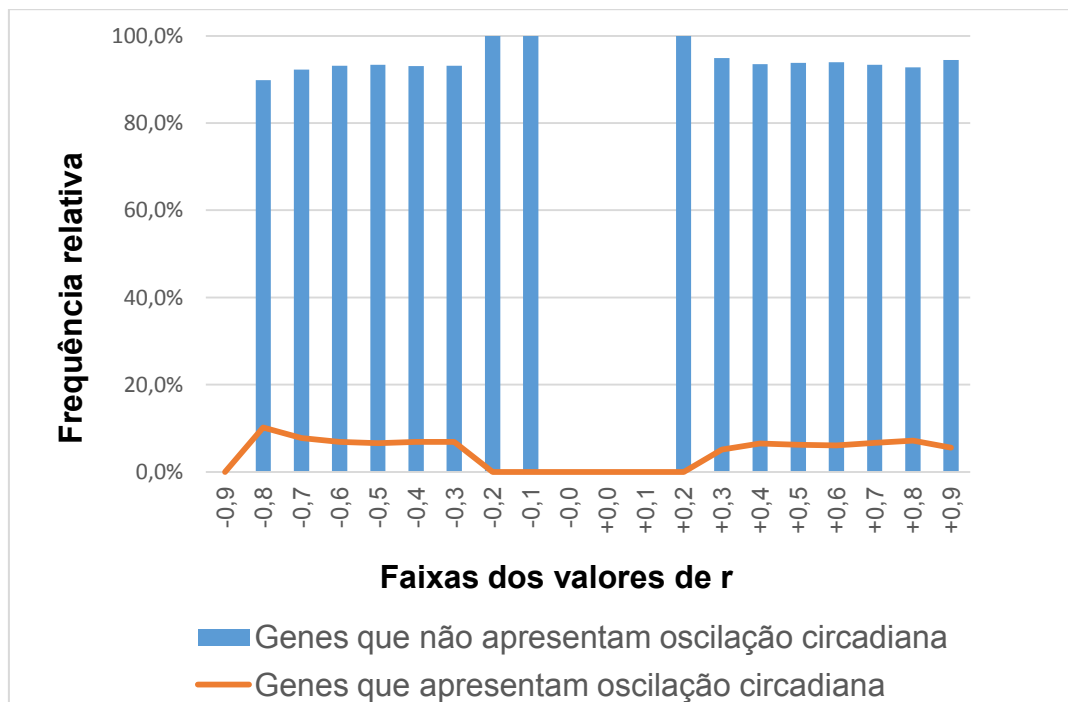
Analisando a distribuição dos dados de correlação considerando estas categorias, ou seja, genes que apresentaram ou não variação circadiana, através da frequência relativa da soma desses dois grupos por faixas de correlações, com intervalos de 0,1, evidenciamos um perfil onde o conjunto de genes que apresentou variação também tende a apresentar maiores valores de correlação, positiva ou negativa, para ambas as espécies (Figuras 14 e 15).

Figura 14. Frequência relativa dos genes correlacionados obtidos para espécie *M. musculus* por faixa dos valores de r , que apresentam ou não oscilação circadiana de acordo com o trabalho de ZANGH et al.,2014.



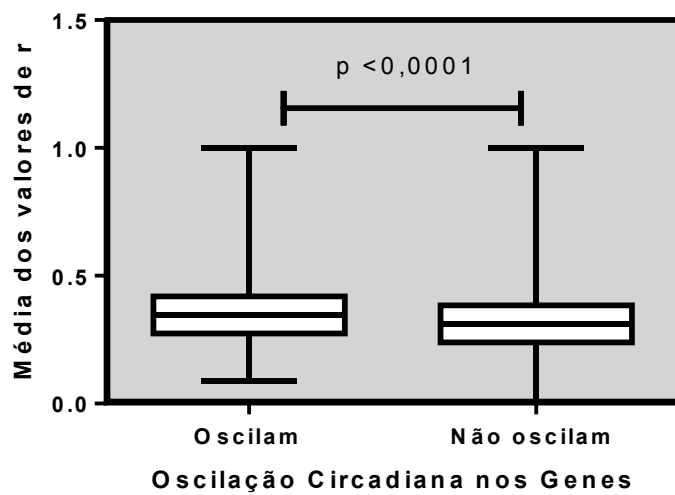
Fonte: Autor

Figura 15. Distribuição em porcentagem dos genes correlacionados obtidos para espécie *H. Sapiens* por faixa dos valores de r , que apresentam ou não oscilação circadiana de acordo com o trabalho de ZANGH et al.,2014.



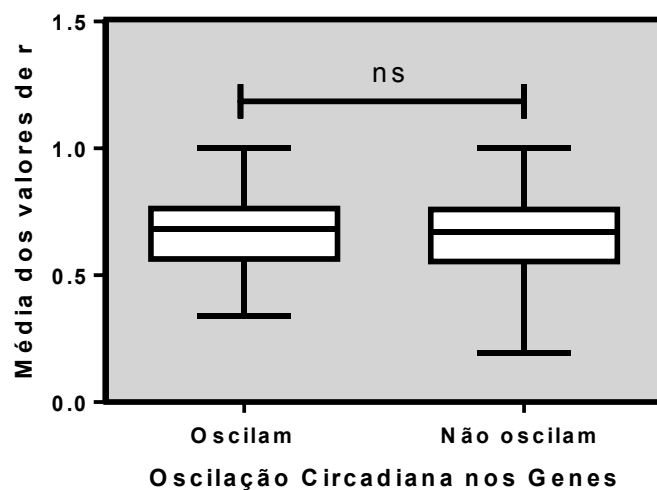
Posteriormente, as médias dos valores *de r* em módulo ($|r|$) destes grupos foram comparadas para ambas as espécies (Figura 16 e 17). Identificamos uma diferença estatisticamente significativa para os valores de $|r|$ em camundongos, onde a média para o grupo que apresenta variação circadiana foi maior ($p < 0,0001$) e não encontramos diferença estatística para os valores de $|r|$ em humanos.

Figura 16. Comparação das médias em módulo dos valores de $|r|$, correspondentes aos genes que, conforme os dados de ZHANG et al.,2014, apresentaram ou não oscilação circadiana em cerebelo de dados obtidos para a espécie *M. musculus* (teste de Mann Whitney $p < 0,0001$).



Fonte: Autor

Figura 17. Comparação das médias em módulo dos valores de $|r|$, correspondentes aos genes que, conforme os dados de LI et al.,2013, apresentaram ou não oscilação circadiana em cerebelo de dados obtidos para a espécie *H. Sapiens* (teste de Mann Whitney $p < 0,0001$).



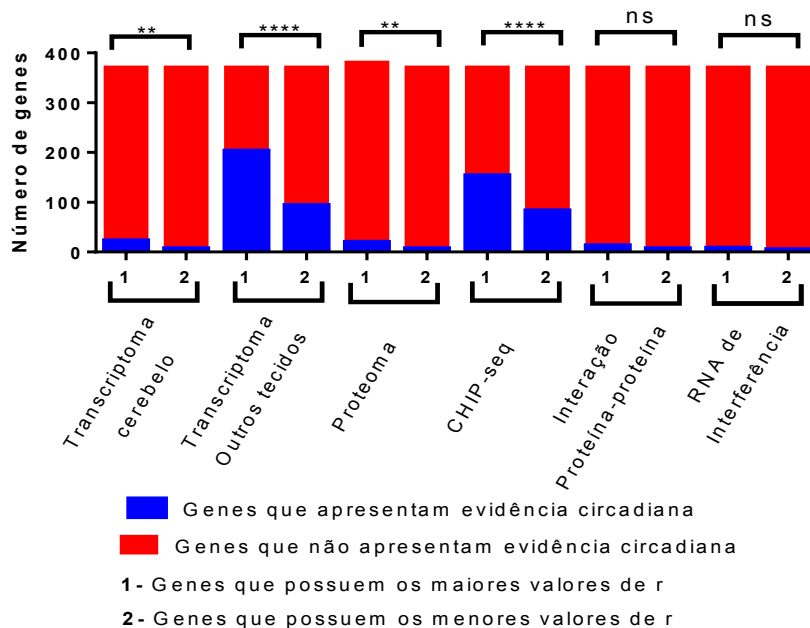
Fonte: Autor

6.4 Análises de outras evidências de envolvimento com o ritmo circadiano.

Com o objetivo de identificar evidências funcionais de interação entre os genes correlacionados e o ritmo circadiano, confrontamos nossos dados com resultados de estudos experimentais em larga escala publicados com ritmos circadianos e envolvendo transcriptoma, proteoma, *ChiP-seq*, interação proteína-proteína e *RNAi*. Seleccionamos 2% dos genes correlacionados que possuem os maiores e os menores valores de *r* positivos e negativos, tanto a para espécie *M. musculus* como para a espécie *H. sapiens*, e analisamos a frequência de evidências destes genes, de acordo com cada estudo experimental.

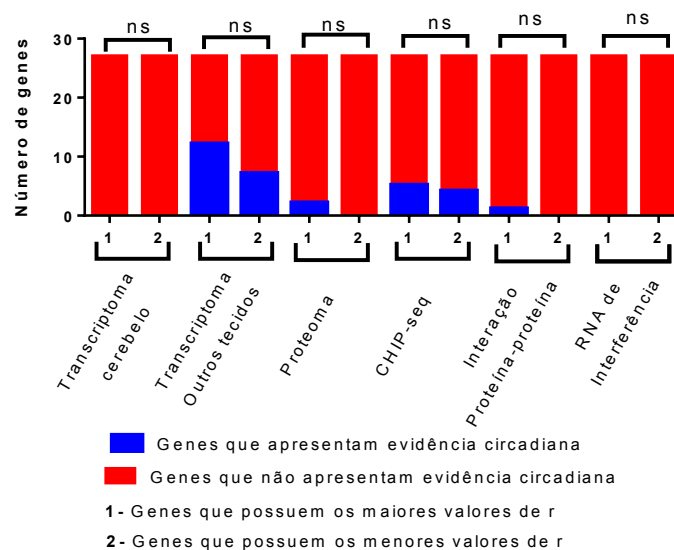
Observamos significância estatística (Qui-quarado $p < 0,01$) em 2% (370) dos genes correlacionados entre os grupos que possuem os maiores valores de *r*, que variam de 0,542 a 0,729 e os menores valores de *r* positivos, entre 0,061 a 0,122, na espécie *Mus músculos*, nos estudos experimentais de transcriptoma cerebelar, outros tecidos, Proteoma e *ChiP-seq*, com uma maior frequência de genes com variação circadiana no grupo de genes com os maiores valores de correlação (Figura 18). Já para os grupos de 2% (27) genes que possuem os maiores valores de *r* negativos, que variam de -0,477 a -0,623 e os menores valores, entre -0,026 a -0,087, não encontramos significância estatística em nenhum dos estudos (Figura 19).

Figura 18. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebelo da espécie *M. musculus*, que possuem os maiores e os menores valores de *r* positivos, nos estudos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e *RNAi* (os asteriscos indicam significância estatística, * para valores de *p* a 0,05, ** para valores inferiores a 0,01 e *** ou mais para valores de *p* inferiores a 0,001; ns - significa não significativa).



Fonte: Autor

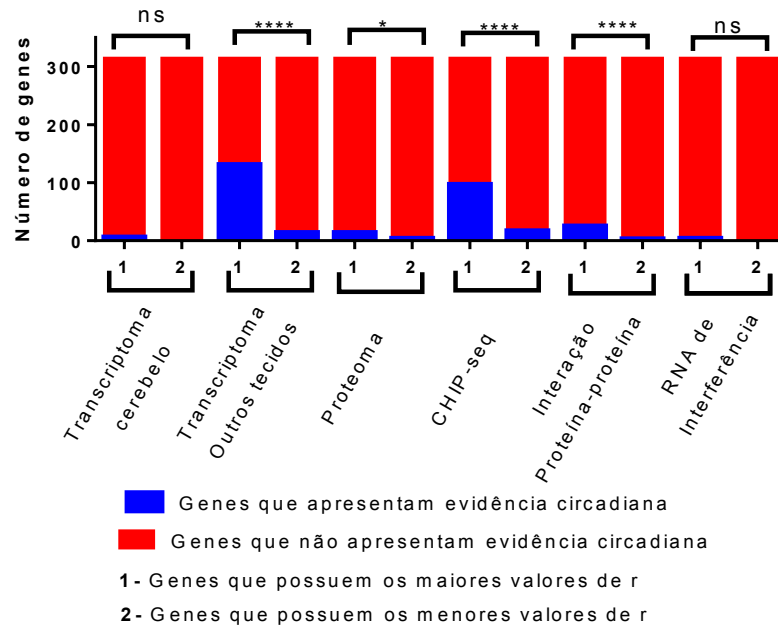
Figura 19. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebello da espécie *M. musculus*, que possuem os maiores e os menores valores de r negativos, nos estudos experimentais de transcriptoma do cerebello e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi (os asteriscos indicam significância estatística, * para valores de p a 0,05, ** para valores inferiores a 0,01 e *** ou mais para valores de p inferiores a 0,001; ns - significa não significante).



Fonte: Autor

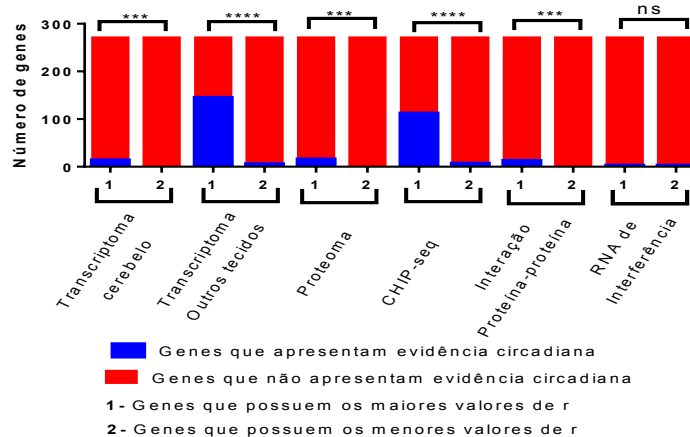
Identificamos na comparação de 2% (313) dos genes correlacionados da espécie *H. sapiens* dos grupos com os maiores e menores valores positivos de r , que variam respectivamente de 0,833 a 0,948 e 0,13 a 0,3, que houve diferença estatística nos estudos experimentais de transcriptoma de outros tecidos, Proteoma, *ChiP-seq* e interação proteína-proteína (Figura 20). Para os grupos de 2% (270) genes com os maiores valores de r negativos, que variam de -0,815 a -0,899 e menores valores de r , que variam de -0,147 a -0,293, observamos diferenças estatísticas em quase todos os estudos experimentais, com exceção do estudo de RNAi (Figura 21). Neste caso também, os grupos com os maiores valores de correlação apresentaram uma maior frequência de genes com evidências de variação circadiana ($p < 0,01$).

Figura 20. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebello da espécie *H. Sapiens*, que possuem os maiores e os menores valores de r positivos, nos estudos experimentais de transcriptoma do cerebello e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi (os asteriscos indicam significância estatística, * para valores de p a 0,05, ** para valores inferiores a 0,01 e *** ou mais para valores de p inferiores a 0,001; ns - significa não significativa).



Fonte: Autor

Figura 21. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de evidências de envolvimento com ritmos circadianos de 2% dos genes correlacionados em cerebello da espécie *H. Sapiens*, que possuem os maiores e os menores valores de r negativos, nos estudos experimentais de transcriptoma do cerebello e outros tecidos e estruturas, estudos de *ChiP-seq*, proteoma, interação proteína-proteína e RNAi (os asteriscos indicam significância estatística, * para valores de p a 0,05, ** para valores inferiores a 0,01 e *** ou mais para valores de p inferiores a 0,001; ns - significa não significativa).



Fonte: Autor

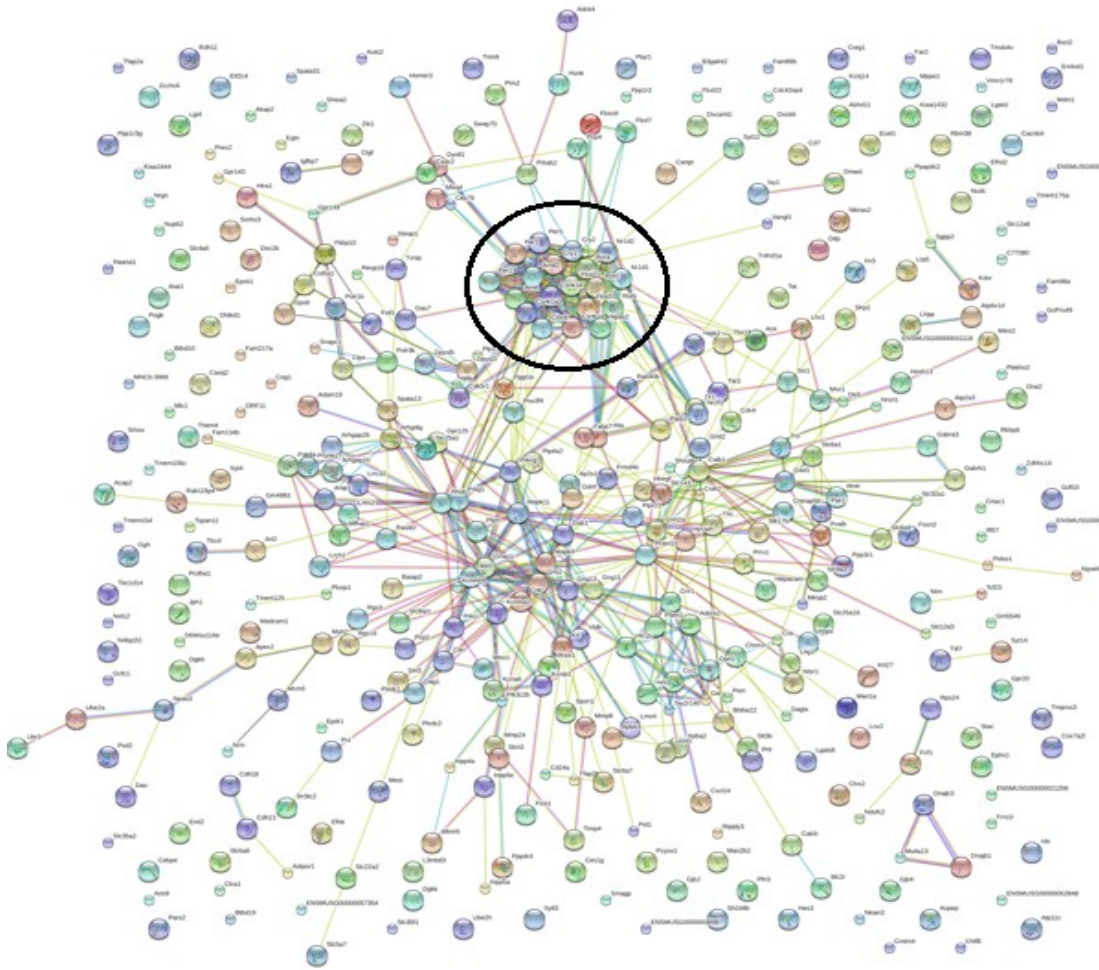
Considerando o conjunto de evidências para *M. musculus*, 259 genes no grupo com maiores valores positivos de correlação apresentaram alguma evidência de interação com o ritmo circadiano, a partir dos estudos em larga escala incluídos na análise, contra 129 genes do grupo com menores valores de correlação e 16 genes com maiores valores negativos, contra 7 genes com menores valores de correlação.

Para *H. sapiens* 160 genes no grupo com maiores valores de correlação positiva e 22 genes no grupo com menores valores, 176 genes no grupo com maiores valores de correlação negativa e 9 genes no grupo com menores valores correlação apresentaram alguma evidência de interação com o ritmo circadiano.

6.5 Redes de interação proteica

Realizamos análises de interações funcionais, através do portal STRING (<http://string-db.org>) das 19 proteínas circadianas estudadas com 397 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas e para as 397 proteínas, correspondentes aos 2% de genes com os menores valores de correlações positivas e negativas, obtidas a partir dos dados de *M. musculus*. Evidenciamos que entre as 416 proteínas analisadas no grupo com maiores valores de correlações negativas e positivas somadas às 19 proteínas circadianas, ocorrem 604 interações funcionais (Figura 22). A maioria das interações estabelece uma rede que se conecta com as proteínas circadianas.

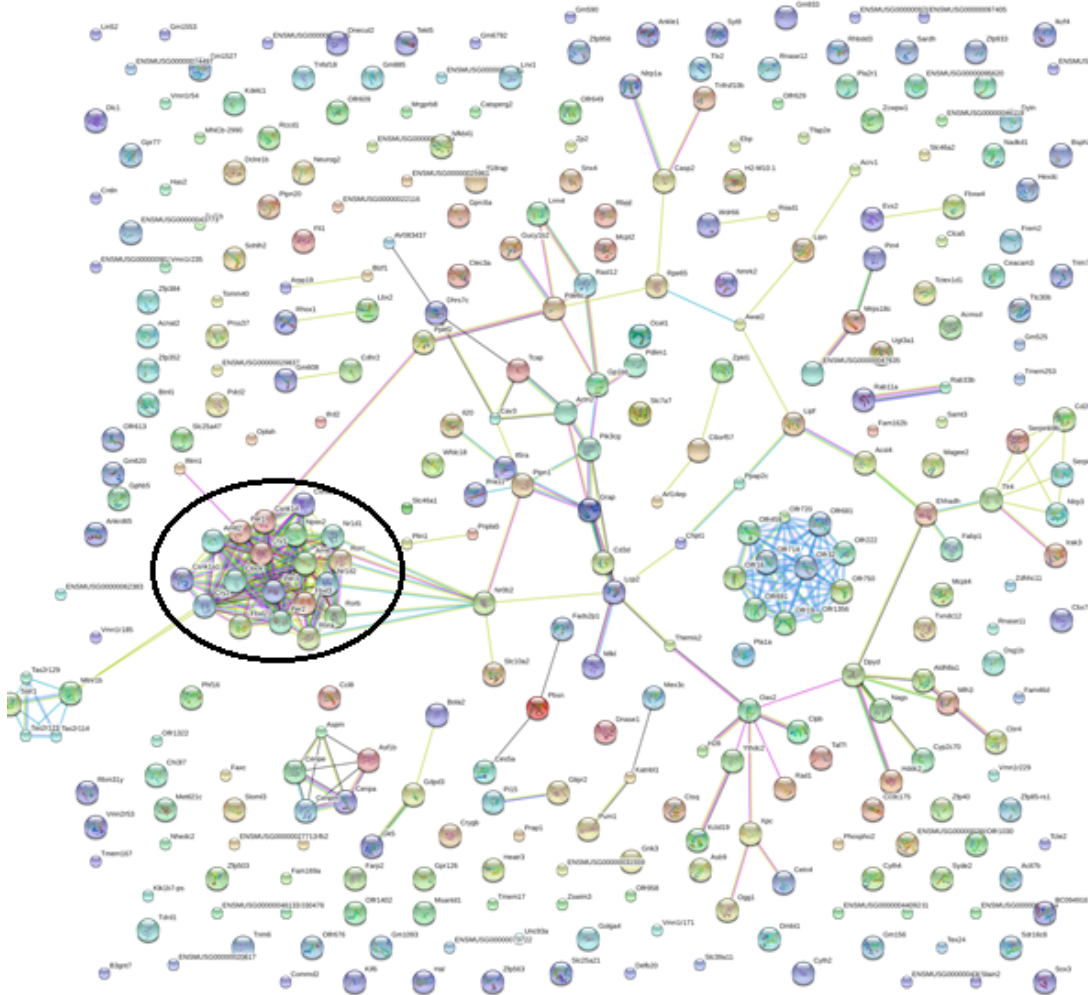
Figura 22. Análise de interações funcionais das 19 proteínas circadianas estudadas com 397 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *M. musculus*, através do portal STRING (<http://string-db.org/>). A área em destaque na figura refere-se a localização das 19 proteínas circadianas estudadas.



Fonte: Autor.

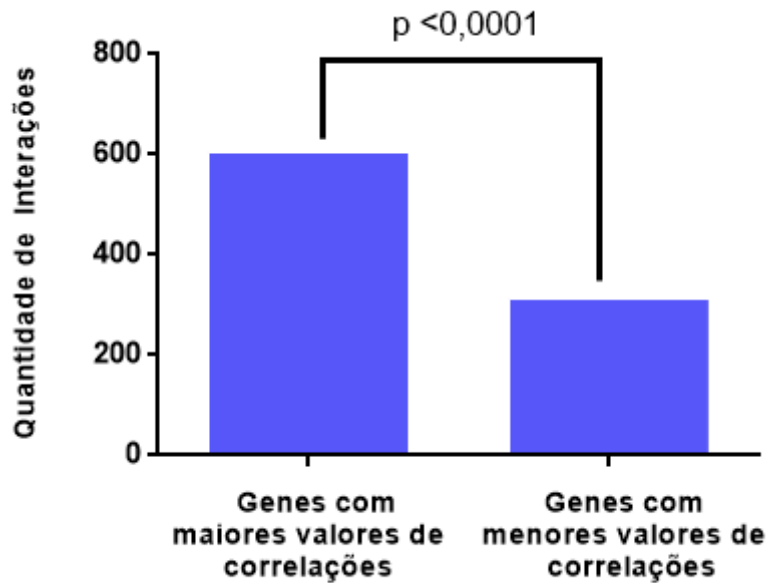
No grupo com 416 proteínas com os menores valores de correlações negativas e positivas, somadas às 19 proteínas circadianas, há 312 interações (Figura 23). Boa parte das interações deste grupo decorre das estabelecidas entre ORGs (*Olfactory Receptor Genes*) que não mantêm relação com as proteínas circadianas. A diferença observada entre os grupos relativa ao número de interações estabelecidas é estatisticamente significativa (Figura 24).

Figura 23. Análise de interações funcionais das 19 proteínas circadianas estudadas com 397 proteínas, correspondentes aos 2% dos genes com os menores valores de correlações positivas e negativas em cérebro de *M. musculus*, através do portal STRING (<http://string-db.org/>). A área em destaque na figura refere-se a localização das 19 proteínas circadianas estudadas.



Fonte: Autor.

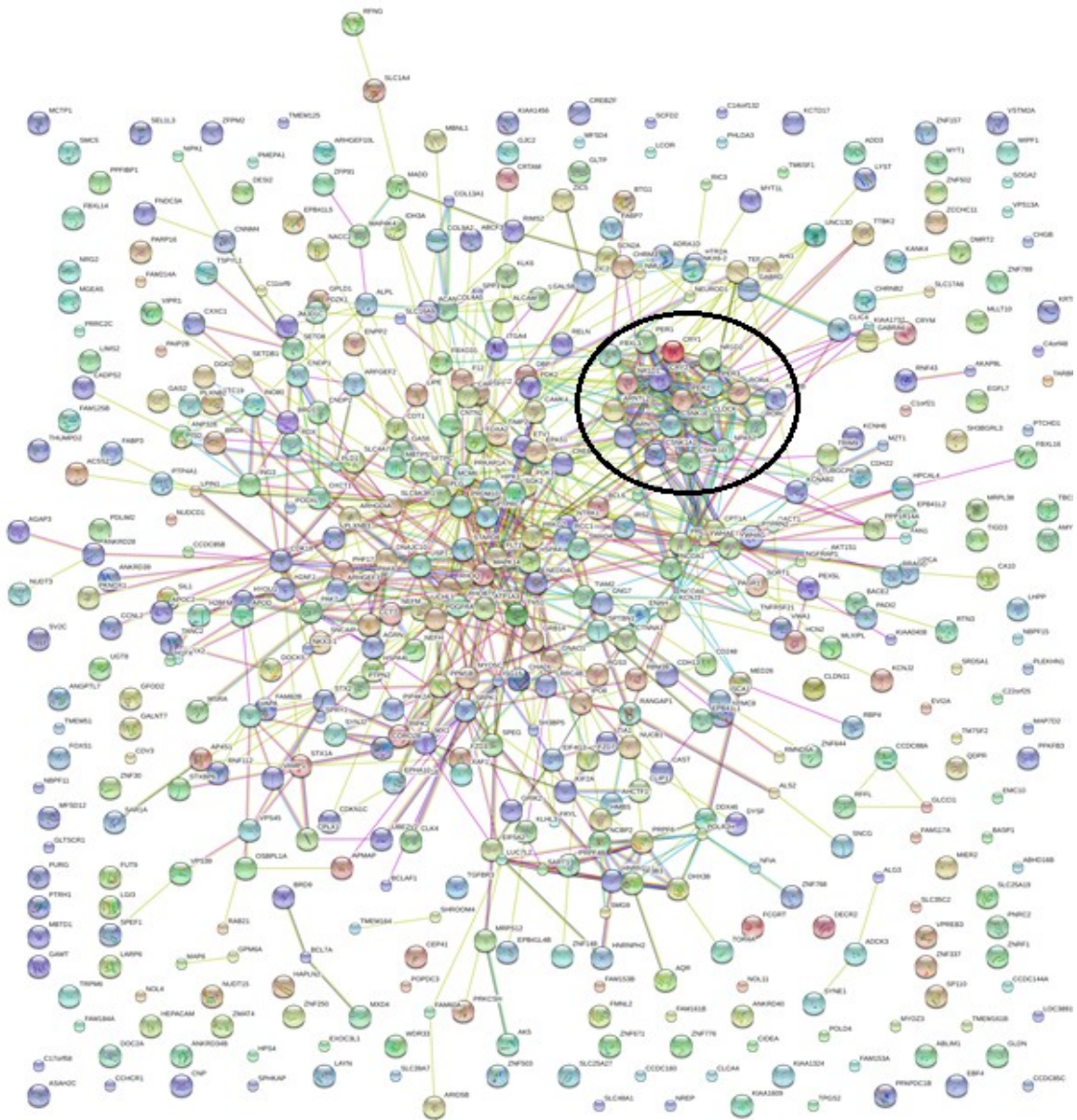
Figura 24. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de interações funcionais das 416 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *M. musculus*, mais as 19 proteínas circadianas e 604 interações, comparados as 416 proteínas com os menores valores de correlações negativos e positivos, somadas às 19 proteínas circadianas, e as 312 interações. Significância estatística com $p < 0,0001$.



Fonte: Autor.

Realizamos também estas mesmas análises de interações funcionais, através do portal STRING, das 19 proteínas circadianas estudadas com 583 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *H. sapiens* e para as 583 proteínas, correspondentes aos 2% de genes com os menores valores de correlações positivas e negativas. Evidenciamos que das 621 proteínas analisadas no grupo com maiores valores de correlações negativos e positivos, há 804 interações funcionais entre elas (Figura 25). Similarmente ao encontrado para os dados de camundongo, a maioria das conexões estabelece uma rede que integra as proteínas circadianas, com exceção da proteína FBXL21, pois não aparece nos resultados para esta espécie.

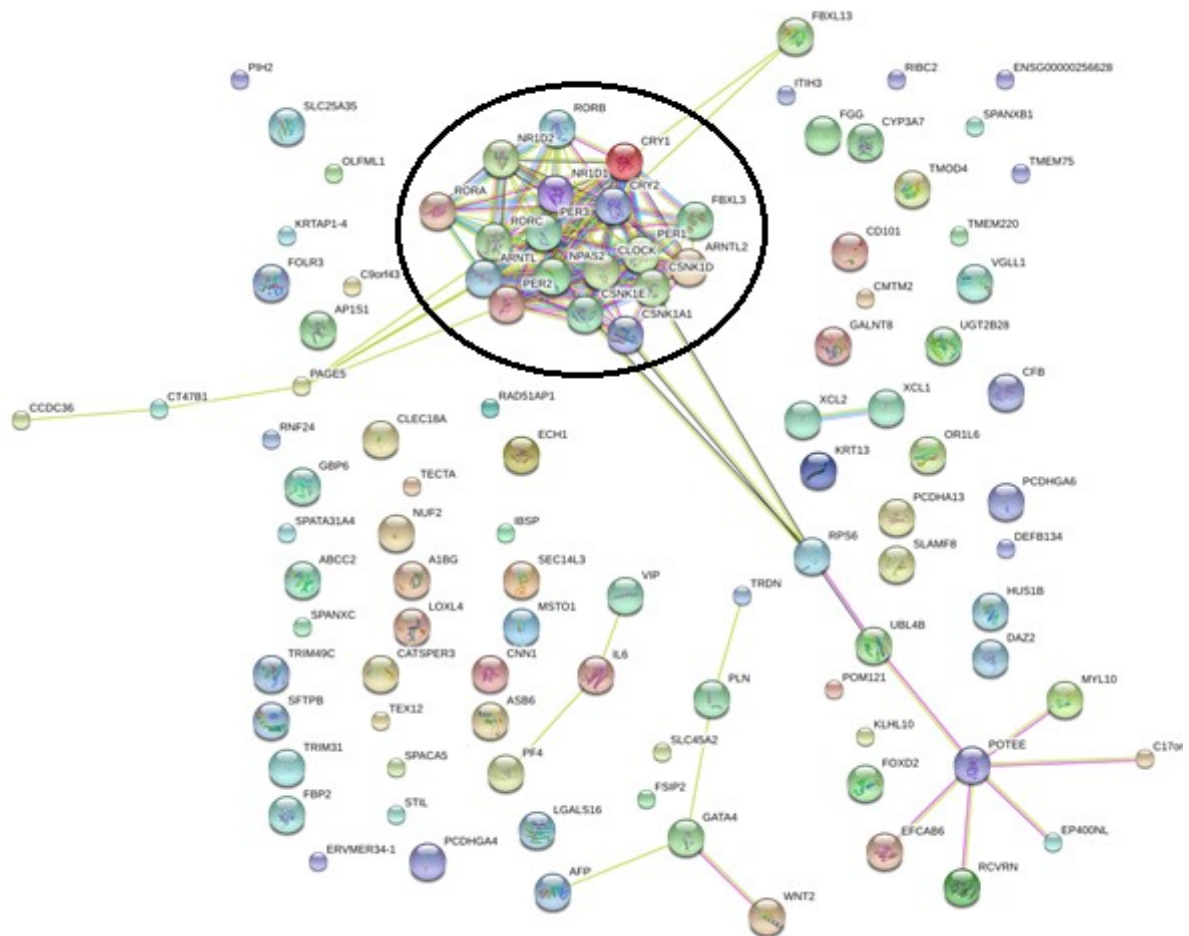
Figura 25. Análise de interações funcionais das 19 proteínas circadianas estudadas com 621 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *H. sapiens*, através do portal STRING (<http://string-db.org/>). A área em destaque na figura refere-se a localização das 19 proteínas circadianas estudadas.



Fonte: Autor.

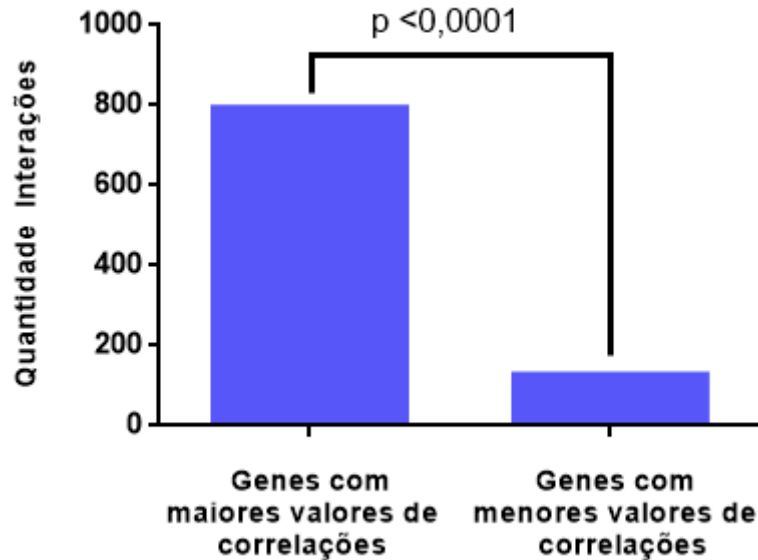
No grupo com 621 proteínas com os menores valores de correlações negativos e positivos, há 138 interações, sendo a maioria delas decorrente das interações entre as proteínas circadianas (Figura 26). Neste caso, a maioria das proteínas não foi descrita nos resultados gerados pelo programa, incluindo a proteína circadiana, FBXL21 (Figura 27).

Figura 26. Análise de interações funcionais das 19 proteínas circadianas estudadas com 621 proteínas, correspondentes aos 2% dos genes com os menores valores de correlações positivas e negativas em cerebelo de *H. sapiens*, através do portal STRING (<http://string-db.org/>). A área em destaque na figura refere-se a localização das 19 proteínas circadianas estudadas.



Fonte: Autor.

Figura 27. Resultado de Qui-quadrado por meio da comparação entre as análises das frequências de interações funcionais das 621 proteínas, correspondentes aos 2% dos genes com os maiores valores de correlações positivas e negativas em cerebelo de *H. sapiens* mais as 19 proteínas circadianas e 804 interações, comparados as 621 proteínas com os menores valores de correlações negativos e positivos, somadas às 19 proteínas circadianas, e as 138 interações. Significância estatística com $p < 0,0001$.



Fonte: Autor.

6.6 Ranqueamento de genes candidatos

Os 100 genes correlacionados com os maiores valores de $|r|$ obtidos para *M. musculus* foram ranqueados com base na soma destes valores com os valores de correlação obtidos para o gene homólogo em *H. sapiens*, a fim de selecionar genes candidatos para futuros ensaios experimentais (Tabela 06). Destes 100 genes correlacionados, 99 possuem valores de r positivos e 1 negativo em camundongos e para humanos, 55 genes possuem valores de r positivos e 45 negativos. Quatro deles são genes circadianos, *Per2* e *Cry2*, com os mesmos valores de r tanto na espécie *M. musculus* quanto em *H. sapiens*, com valor de 0,603 para *Per2* e 0,855 e para *Cry2*. *Bmal1* e *Rora* apresentaram valores de r de 0,628 e 0,8 e 0,628 e 0,651 para *M. musculus* e *H. sapiens*, respectivamente. *Per2* e *Cry2* têm correlações recíprocas nas duas espécies. Ademais, observamos que 2 genes (*Slc1a3* e *Rora*) foram correlacionados com o gene circadiano *Bmal1* tanto em camundongo como em humanos, 03 genes (*Calb1*, *Gdf10* e *Stk17b*) e 01 (*Msh2*), foram correlacionados com *Rora* e *Csnkla*,

respectivamente. Além destes, nenhum outro gene apresentou correlação com o mesmo gene relógio em ambas as espécies.

Tabela 06. 100 genes correlacionados com os maiores valores de $|r|$ obtidos para *M. musculus* ranqueados com base na soma dos valores de correlação obtidos para o gene homólogo em *H. sapiens*, além de dados experimentais de estudos cronobiológicos.

NOME DO GENE	CORRELAÇÃO (r)				TRANSCRIPTOMA		PROTEOMA	CHIP-SEQ	PROTEÍN A - PROTEÍNA	RNAI	SOMA DOS VALORES DE R	ESCORE DE EVIDÊNCIA
	Camundongo	Gene Circadiano	Humano	Gene Circadiano	Cerebelo	Outros Tecidos	Diferentes Tecidos	Diferentes Genes Circadianos				
<i>Cartpt</i>	0,729	<i>Rorb</i>	-0,827	<i>Cry2</i>	0	0	0	0	0	0	1,556	0
<i>Gpr149</i>	0,725	<i>Npas2</i>	-0,814	<i>Cry2</i>	0	0	0	0	0	0	1,539	0
<i>Ptpr</i>	0,667	<i>Rora</i>	0,863	<i>Cry2</i>	0	0	0	0	0	0	1,53	0
<i>Cnr1</i>	0,665	<i>Rora</i>	0,856	<i>Per3</i>	0	1	0	0	0	0	1,521	1
<i>Inpp4a</i>	0,721	<i>Bmal1</i>	0,797	<i>Cry2</i>	0	0	0	1	0	0	1,518	1
<i>Prdm10</i>	0,623	<i>Npas2</i>	0,894	<i>Cry2</i>	0	4	0	0	0	0	1,517	4
<i>Gnao1</i>	0,657	<i>Bmal1</i>	-0,827	<i>Per2</i>	0	1	1	0	0	0	1,484	2
<i>Baiap2</i>	0,604	<i>Rora</i>	0,873	<i>Cry2</i>	0	7	0	4	0	0	1,477	11
<i>Grid2</i>	0,678	<i>Rora</i>	0,784	<i>Cry2</i>	0	0	0	0	0	0	1,462	0
<i>Nipal4</i>	0,673	<i>Npas2</i>	-0,787	<i>Cry2</i>	0	0	0	0	0	0	1,46	0
<i>Cry2</i>	0,603	<i>Per2</i>	0,855	<i>Per2</i>	1	10	0	7	1	1	1,458	20
<i>Per2</i>	0,603	<i>Cry2</i>	0,855	<i>CRY2</i>	1	15	0	8	1	0	1,458	25
<i>Cdh13</i>	0,609	<i>Rorb</i>	-0,846	<i>Cry2</i>	0	2	0	0	0	0	1,455	2
<i>Cdc42ep4</i>	0,626	<i>Rora</i>	0,825	<i>Cry2</i>	1	5	0	6	0	0	1,451	12
<i>Ecel1</i>	0,663	<i>Rorb</i>	-0,777	<i>Cry2</i>	0	0	0	0	0	0	1,44	0
<i>Jph1</i>	0,609	<i>Fbxl21</i>	0,829	<i>Cry2</i>	0	2	0	1	0	0	1,438	3
<i>Syt4</i>	0,634	<i>Nr1d1</i>	0,8	<i>Cry1</i>	0	0	0	0	0	0	1,434	0
<i>Timp4</i>	0,63	<i>Rora</i>	0,804	<i>Cry2</i>	0	2	0	0	0	0	1,434	2
<i>Prune2</i>	0,635	<i>Rorb</i>	-0,793	<i>Bmal1</i>	0	12	0	0	0	0	1,428	12
<i>Bmal1</i>	0,628	<i>Rora</i>	0,8	<i>Per3</i>	1	4	0	2	1	0	1,428	8
<i>Rab11fip4</i>	0,649	<i>Bmal1</i>	0,778	<i>Csnk1d</i>	0	1	0	0	0	0	1,427	1
<i>Gcfc1</i>	0,635	<i>Rora</i>	0,786	<i>Nr1d2</i>	0	0	0	3	1	0	1,421	4
<i>Abi1</i>	0,613	<i>Npas2</i>	-0,805	<i>Cry2</i>	0	2	0	7	0	1	1,418	10

<i>Calb1</i>	0,676	<i>Rora</i>	0,736	<i>Rora</i>	0	2	0	0	0	0	1,412	2
<i>Mtnr6</i>	0,675	<i>Csnk1a</i>	-0,737	<i>Nr1d1</i>	0	1	0	2	0	0	1,412	3
<i>Atp2a3</i>	0,639	<i>Rora</i>	0,773	<i>Bmal1</i>	0	2	0	0	0	0	1,412	2
<i>Msh2</i>	0,615	<i>Csnk1a</i>	0,793	<i>Csnk1a</i>	0	0	0	1	0	0	1,408	1
<i>Nrcam</i>	0,621	<i>Rora</i>	-0,784	<i>Per1</i>	0	1	0	0	0	0	1,405	1
<i>Rasal2</i>	0,612	<i>Rora</i>	0,793	<i>Csnk1d</i>	0	3	0	0	0	0	1,405	3
<i>Tmem106c</i>	0,599	<i>Csnk1e</i>	0,806	<i>Bmal1</i>	0	1	0	0	0	0	1,405	1
<i>Csdc2</i>	0,648	<i>Rora</i>	0,752	<i>Bmal1</i>	0	2	0	0	0	0	1,4	2
<i>Slc6a6</i>	0,645	<i>Bmal1</i>	-0,754	<i>Per2</i>	0	7	0	6	0	0	1,399	13
<i>Pggt1b</i>	0,698	<i>Fbxl21</i>	0,694	<i>Csnk1d</i>	0	0	0	0	0	0	1,392	0
<i>Epsii1</i>	0,677	<i>Npas2</i>	-0,715	<i>Bmal1</i>	0	0	0	0	0	0	1,392	0
<i>Apex2</i>	0,632	<i>Npas2</i>	-0,753	<i>Csnk1a</i>	0	2	0	0	0	0	1,385	2
<i>Lnx2</i>	0,599	<i>Fbxl21</i>	0,778	<i>Cry2</i>	0	3	0	3	0	0	1,377	6
<i>Cep76</i>	0,634	<i>Rora</i>	0,741	<i>Bmal1</i>	0	2	0	0	0	0	1,375	2
<i>Sgpp2</i>	0,604	<i>Npas2</i>	0,768	<i>Csnk1d</i>	1	2	0	0	0	0	1,372	3
<i>Slc1a3</i>	0,622	<i>Bmal1</i>	0,745	<i>Bmal1</i>	0	0	1	1	0	0	1,367	2
<i>Ppp3r1</i>	-0,623	<i>Rorb</i>	0,744	<i>Cry1</i>	0	5	0	4	0	0	1,367	9
<i>Endod1</i>	0,6	<i>Csnk1e</i>	0,764	<i>Csnk1d</i>	0	1	0	3	0	0	1,364	4
<i>Gdf10</i>	0,639	<i>Rora</i>	0,724	<i>Rora</i>	1	5	0	0	0	0	1,363	6
<i>Prex1</i>	0,625	<i>Rora</i>	-0,734	<i>Cry2</i>	0	1	0	3	0	0	1,359	4
<i>Auts2</i>	0,657	<i>Rora</i>	-0,696	<i>Bmal1</i>	0	2	0	0	0	0	1,353	2
<i>Pcyox1</i>	0,61	<i>Csnk1e</i>	0,741	<i>Csnk1a</i>	0	3	0	7	1	0	1,351	11
<i>Stk17b</i>	0,628	<i>Rora</i>	0,722	<i>Rora</i>	0	2	0	0	0	0	1,35	2
<i>Atp6v1d</i>	0,6	<i>Csnk1a</i>	-0,746	<i>Nr1d1</i>	0	6	0	0	0	0	1,346	6
<i>Homer3</i>	0,606	<i>Bmal1</i>	0,723	<i>Cry2</i>	0	0	0	0	0	1	1,329	1
<i>Cog1</i>	0,614	<i>Fbxl21</i>	0,714	<i>Csnk1e</i>	0	0	0	0	0	0	1,328	0
<i>Arl2</i>	0,606	<i>Csnk1a</i>	-0,715	<i>Cry1</i>	0	0	0	1	0	0	1,321	1
<i>Syt12</i>	0,602	<i>Rora</i>	0,718	<i>Cry2</i>	1	3	0	0	0	0	1,32	4
<i>Dscam1l</i>	0,604	<i>Per3</i>	-0,711	<i>Cry2</i>	0	0	0	0	0	0	1,315	0

Hbegf	0,637	<i>Bmall</i>	0,672	<i>Per1</i>	0	1	0	3	0	0	1,309	4
Spata13	0,647	<i>Fbxl21</i>	-0,661	<i>Cry2</i>	0	2	0	6	0	0	1,308	8
Slc5a7	0,711	<i>Rorb</i>	-0,591	<i>Per2</i>	0	0	0	0	0	0	1,302	0
Pde9a	0,638	<i>Bmall</i>	0,648	<i>Cry2</i>	1	3	0	6	0	0	1,286	10
Rab40b	0,623	<i>Per3</i>	0,662	<i>Csnk1d</i>	0	1	0	2	0	0	1,285	3
Prkg1	0,611	<i>Rora</i>	0,671	<i>Rorc</i>	1	5	0	0	0	0	1,282	6
Pcp2	0,651	<i>Rora</i>	-0,628	<i>Cry1</i>	0	0	0	0	0	0	1,279	0
Rora	0,628	<i>Bmall</i>	0,651	<i>Bmall</i>	0	2	0	0	1	0	1,279	3
Ace	0,644	<i>Csnk1d</i>	-0,632	<i>Cry1</i>	0	1	0	0	0	0	1,276	1
Rdh11	0,651	<i>Rora</i>	-0,621	<i>Nr1d1</i>	0	2	1	3	0	0	1,272	6
Arap1	0,613	<i>Bmall</i>	-0,648	<i>Cry1</i>	0	3	0	4	0	0	1,261	7
Creg1	0,661	<i>Rora</i>	0,598	<i>Csnk1d</i>	0	2	0	6	0	0	1,259	8
Parp1	0,646	<i>Rora</i>	-0,613	<i>Cry2</i>	0	0	1	2	1	0	1,259	4
Ntsr1	0,616	<i>Npas2</i>	-0,641	<i>Per2</i>	0	0	0	0	0	0	1,257	0
Gal	0,713	<i>Rorb</i>	-0,542	<i>Nr1d1</i>	0	0	0	0	0	0	1,255	0
Shisa6	0,702	<i>Rora</i>	-0,553	<i>Per1</i>	0	1	0	0	0	0	1,255	1
Tat	0,652	<i>Nr1d1</i>	-0,596	<i>Cry1</i>	0	1	2	5	0	0	1,248	8
Prrx1	0,607	<i>Csnk1d</i>	0,639	<i>Rorb</i>	0	0	0	2	0	0	1,246	2
Lgi4	0,643	<i>Rora</i>	0,593	<i>Rorc</i>	1	2	0	0	0	0	1,236	3
Htr1b	0,646	<i>Bmall</i>	-0,589	<i>Per2</i>	0	1	0	0	0	0	1,235	1
Gng13	0,684	<i>Rora</i>	0,544	<i>Cry2</i>	0	0	0	0	0	0	1,228	0
Kenma1	0,621	<i>Rora</i>	-0,605	<i>Per2</i>	0	1	0	0	0	0	1,226	1
Slc35a2	0,604	<i>Nr1d1</i>	-0,621	<i>Cry2</i>	0	1	0	4	0	0	1,225	5
Mapk11	0,602	<i>Rorb</i>	0,619	<i>Clock</i>	0	0	0	0	0	0	1,221	0
Prkcg	0,615	<i>Rora</i>	0,601	<i>Csnk1e</i>	0	2	0	0	1	0	1,216	3
Tmem154	0,638	<i>Fbxl21</i>	-0,577	<i>Per2</i>	0	0	0	2	0	0	1,215	2
Hrh3	0,614	<i>Npas2</i>	0,587	<i>Per2</i>	0	1	0	0	0	0	1,201	1
Foxi1	0,63	<i>Nr1d1</i>	-0,565	<i>Per3</i>	0	0	0	0	0	0	1,195	0
Gjb4	0,602	<i>Rorb</i>	-0,572	<i>Nr1d2</i>	0	0	0	0	0	0	1,174	0

Dagla	0,612	<i>Rora</i>	0,559	<i>Cry1</i>	0	0	0	0	0	0	1,171	0
Nrgn	0,633	<i>Rora</i>	-0,531	<i>Per2</i>	0	0	0	0	0	0	1,164	0
Slc6a9	0,644	<i>Bmal1</i>	-0,512	<i>Cry1</i>	0	2	0	7	0	0	1,156	9
Slc4a5	0,602	<i>Csnk1d</i>	0,554	<i>Per2</i>	1	3	0	0	0	0	1,156	4
Efhb	0,626	<i>Rora</i>	-0,525	<i>Rorc</i>	0	0	0	0	0	0	1,151	0
Fam134b	0,614	<i>Csnk1a</i>	-0,529	<i>Csnk1e</i>	0	2	0	6	0	0	1,143	8
Sh2d4b	0,632	<i>Fbxl21</i>	-0,51	<i>Nr1d2</i>	0	2	0	1	0	0	1,142	3
Rgs16	0,615	<i>Fbxl21</i>	0,527	<i>Rora</i>	1	1	0	8	0	0	1,142	10
Prlr	0,614	<i>Npas2</i>	-0,527	<i>Nr1d2</i>	1	5	0	6	0	0	1,141	12
Tbx19	0,6	<i>Fbxl21</i>	0,533	<i>Rorc</i>	0	0	0	0	0	0	1,133	0
Dhtkd1	0,606	<i>Fbxl21</i>	0,521	<i>Fbxl3</i>	0	3	0	2	0	0	1,127	5
Hes3	0,609	<i>Rora</i>	-0,51	<i>Nr1d2</i>	0	0	0	0	0	0	1,119	0
Ano9	0,669	<i>Npas2</i>	-0,447	<i>Nr1d2</i>	0	0	0	1	0	0	1,116	1
L3mbtl3	0,636	<i>Npas2</i>	0,48	<i>Csnk1a</i>	0	7	0	1	0	0	1,116	8
Pou3f4	0,617	<i>Fbxl21</i>	0,496	<i>Rorb</i>	0	0	0	0	0	0	1,113	0
Gpr20	0,618	<i>Npas2</i>	0,485	<i>Rorc</i>	0	0	0	0	0	0	1,103	0
Fkbp10	0,633	<i>Nr1d1</i>	-0,464	<i>Nr1d2</i>	0	4	0	0	0	0	1,097	4
Cdx4	0,65	<i>Per3</i>	-0,445	<i>Nr1d2</i>	0	1	0	2	0	0	1,095	3
Pars2	0,64	<i>Fbxl21</i>	0,455	<i>Nr1d1</i>	0	0	0	0	0	0	1,095	0

Fonte: Autor

O gene *Cartpt* possui o maior valor r positivo encontrado em *M. musculus* de 0,729, correlacionado com o gene circadiano *Rorb*, e um valor de r de -0,827, encontrado em *H. Sapiens*, correlacionado negativamente com *Cry2*, ocupando a primeira posição da tabela, apresentando o maior resultado da somatória entre os valores de $|r|$ obtidos para duas espécies. O gene *Ppp3r1* possui o maior valor negativo encontrado em *M. musculus* de 0,623, e também está correlacionado com o gene circadiano *Rorb*, e um valor de r de 0,744 encontrado em *H. Sapiens*, correlacionado positivamente com *Cry1*, ocupando a 40ª posição da tabela, de acordo com a somatória entre os valores de $|r|$ obtidos para duas espécies. Comparando o padrão de expressão dos mesmos com o padrão de expressão do gene *Rorb* no cerebelo de *M. musculus*, e com o padrão de expressão do gene *Cry2* e *Cry1* em cerebelo de *H. Sapiens*, utilizando imagens de expressão do cérebro e cerebelo através da técnica de *HIS*, e *microarray*, disponíveis no portal *Allan Brain Atlas* (Figuras 28 e 29) evidenciamos através dessas imagens essa correlação espacial de expressão, positiva e negativa, conforme foi descrito nos valores de r citados acima.

Figura 28. Imagens obtidas através do portal *Allen Brain Atlas*, do padrão de expressão espacial dos genes *Rorb*, *Cartp* e *Ppp3r1* através da técnica de *HIS* no cérebro de *M. musculus*. As imagens (A, C e E), correspondem respectivamente à marcação do RNA no tecido, dos genes *Rorb*, *Cartp* e *Ppp3r1*. As imagens (B, D e F), mostram apenas a expressão dos genes *Rorb*, *Cartp* e *Ppp3r1*. O quadrante destacado nas imagens corresponde à região do cerebelo. O gene *Cartpt* possui correlação alta positiva, valor de 0,729 e o gene *Ppp3r1* possui correlação alta negativa, valor de -0,623 com *Rorb*.

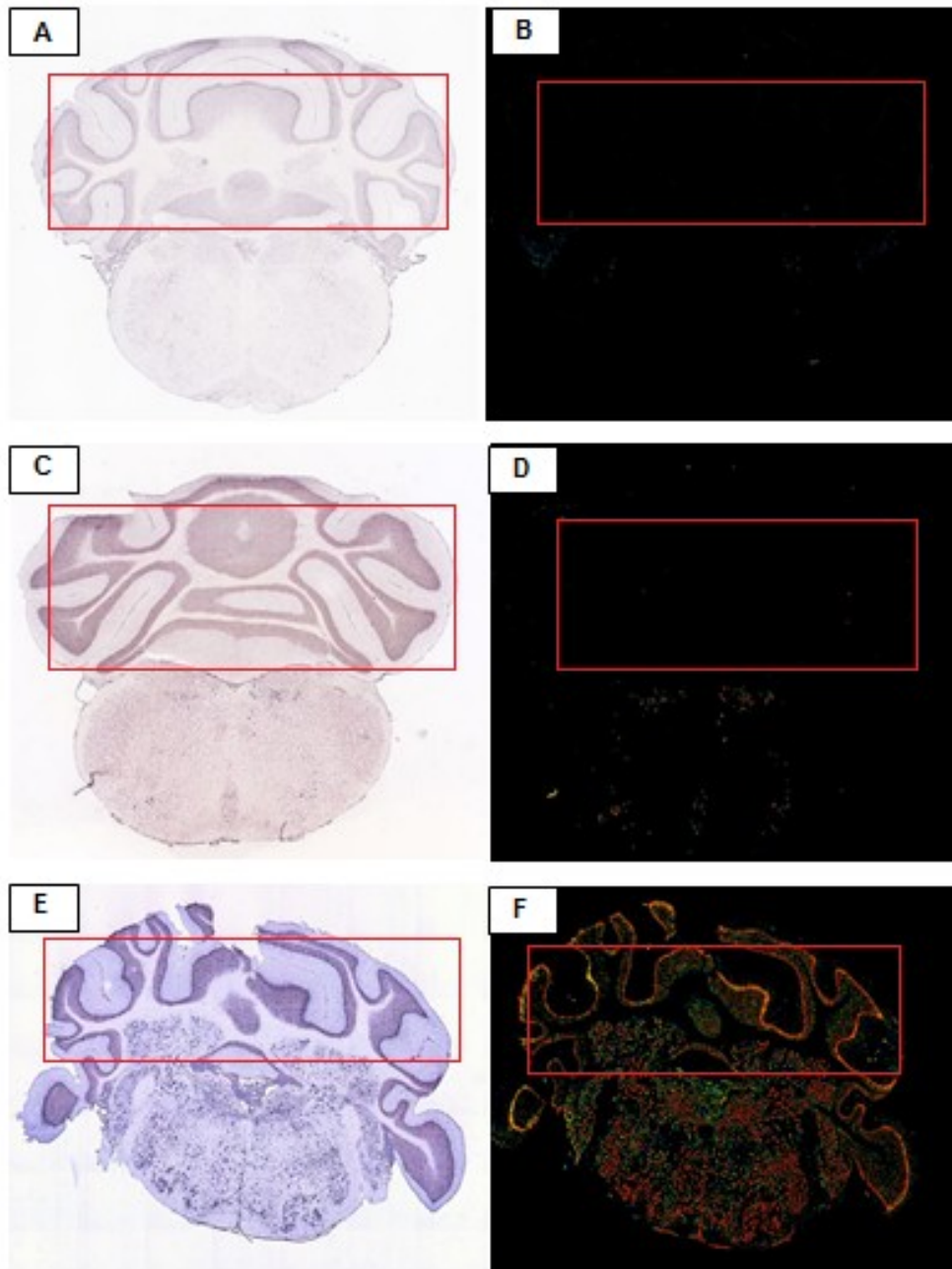
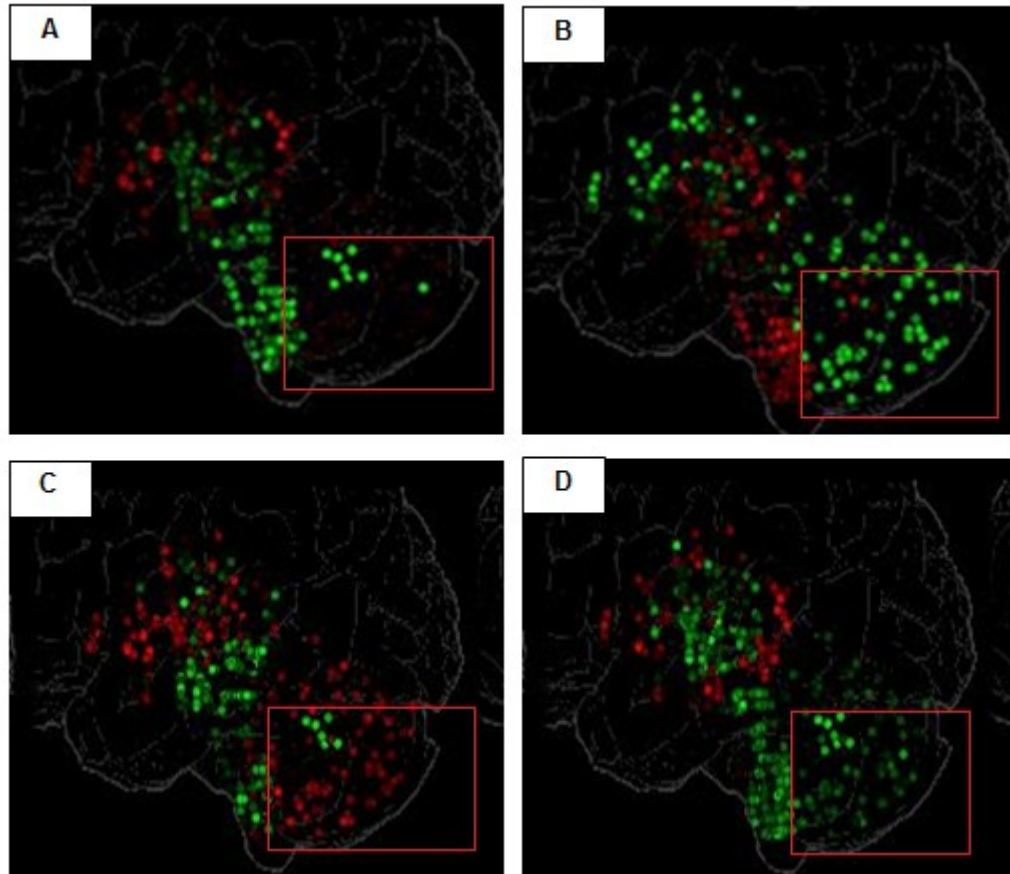


Figura 29. Imagens obtidas através do portal *Allen Brain Atlas*, do padrão de expressão espacial dos genes *Cry2*, *Cartp*, *Cry1* e *Ppp3r1* através da técnica de *Microarray* no cérebro de *H. Sapiens*. As imagens (A e B) correspondem, respectivamente, a expressão dos genes *Cry2* e *Cartpt*. As imagens (C e D), a expressão dos genes *Cry1* e *Ppp3r1*, respectivamente. O quadrante destacado nas imagens corresponde à região do cerebelo e os pontos coloridos indicam a expressão dos genes. Quanto mais próximo do vermelho, maior a expressão, e mais próximo do verde, menor a expressão. O gene *Cartpt* possui correlação alta negativa com *Cry2*, valor de -0,827 e o gene *Ppp3r1* possui correlação alta positiva com *Cry1*, com valor de 0,744.



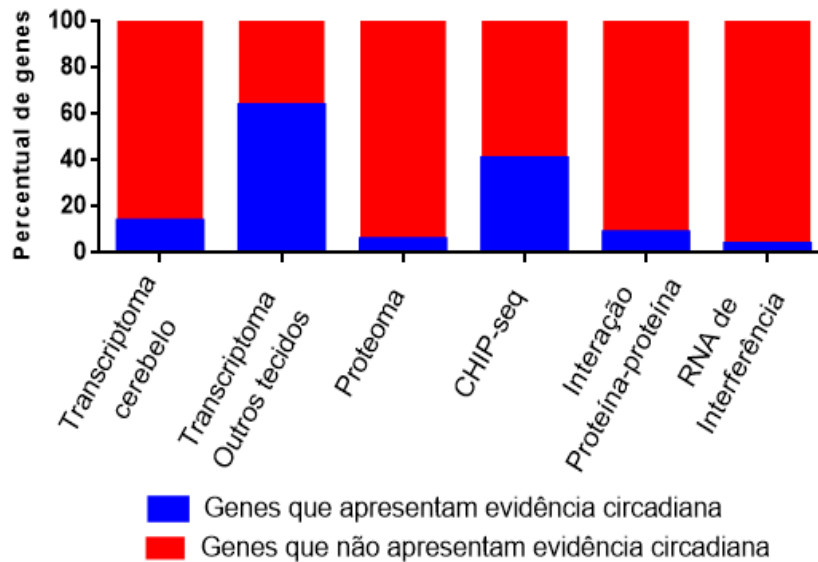
Observou-se também que os genes: *Per2*, *Cry2*, *Slc6a6*, *Cdc42ep4*, *Prune2*, *Prlr*, *Baiap2* e *Pcyox1*, apresentaram os maiores escores acima de 10, resultantes da somatória dos valores de evidências nos diferentes estudos circadianos experimentais. Além disso, dentre os 100 genes, alguns pertencem à mesma família, como os genes *Gcfc1* e *Gdf10*; *Prkcg* e *Prkg1*; *Slc1a3*, *Slc35a2*, *Slc4a5*, *Slc5a7* e *Slc6a6*.

Considerando o estudo de transcriptoma circadiano cerebelar, 13% destes genes apresentam oscilação circadiana e, de acordo com os dados coletados a partir de estudos de transcriptoma em outras estruturas e tecidos, 63% apresentam oscilação circadiana, (Figura 30). Quando confrontamos esses genes com estudos de *ChiP-seq*, identificamos que 40% possuem sítios de ligações para uma ou mais proteínas circadiana.

Para os estudos com proteoma, encontramos que 95 dos 100 genes têm proteínas codificadas que não oscilam circadianamente e que 5 (GNAO1, SLC1A3, RDH11, PARP1 e TAT) possuem oscilação circadiana. Na análise de interação proteína-proteína, 8% das proteínas codificadas por estes 100 genes apresentam interação com alguma proteína circadiana. Três destes genes (*Cry2*, *Abl1* e *Homer3*) após sofrerem *knockdown* por meio de RNAi, induziram mudanças de amplitude e período no ritmo bioluminescente da expressão do gene repórter de luciferase (BMAL1-LUC) em cultura de células.

No conjunto total, 73 genes apresentaram alguma evidência de interação com o ritmo circadiano, previamente identificada nos estudos em larga escala incluídos nesta pesquisa.

Figura 30. Percentual de evidências de 100 genes candidatos identificados com base na correlação de expressão espacial no cérebro de *M. musculus* e *H. sapiens*, a partir de dados disponíveis no Allen Brain Atlas, nos estudos circadianos experimentais de transcriptoma do cerebelo e outros tecidos e estruturas, estudos de *ChIP-seq*, proteoma, interação proteína-proteína e RNAi.



Fonte: Autor

7. DISCUSSÃO

Diferentes abordagens e métodos têm sido utilizados na identificação de genes envolvidos com o ritmo circadiano. Uma abordagem mais atual, conhecida como mineração de dados, surgida a partir da bioinformática, tem sido utilizada em diferentes áreas, inclusive na cronobiologia (LOPES, Robson da Silva et al.2013). O gene *Chrono*, por exemplo, foi identificado dentre 1000 genes candidatos como um importante componente do relógio circadiano em mamíferos, por meio de uma abordagem probabilística computacional, integrando bancos de dados heterogêneos com informações de larga escala de transcriptoma circadiano, expressão gênica em diferentes tecidos, filogenia de genes, vias de interações de proteínas e RNAi. Através de dois testes experimentais, teve seu envolvimento comprovado com BMAL1, interrompendo sua atividade transcricional. Além disto, camundongos que não possuem esse gene apresentaram mudanças nos ritmos da atividade locomotora (ANAFI, Ron C. et al.,2014).

Em outro estudo foram utilizados os 1.000 genes candidatos a reguladores dos ritmos circadianos, obtidos do estudo ANAFI e colaboradores (2014), para uma análise de mineração de dados de larga escala com outros estudos circadianos de proteoma, *ChiP-seq* e interação proteína-proteína, demonstrando 11 genes candidatos, que podem fazer parte do sistema molecular de temporização de mamíferos, propondo seis destes genes, que estão envolvidos com metabolismo de câncer, para investigações experimentais (BHARGAVA, Anuprabha; HERZEL, Hanspeter; ANANTHASUBRAMANIAM, Bharath.,2015).

LEHMANN e colaboradores (2015) desenvolveram um *pipeline* de bioinformática que integra diferentes fontes e tipos de informações de mineração de dados coletados no portal *Pubmed*. Além disso, análises de co-expressão também foram utilizadas para comprovar o envolvimento de 118 potenciais genes controlados pelo relógio estão envolvidos com os ritmos circadianos em mamíferos. (LEHMANN, Robert et al.,2015).

Neste trabalho, identificamos genes que apresentam padrão de expressão espacial no cerebelo humano e murino correlacionada a genes circadianos bem caracterizados. A busca de genes comum com padrão de expressão espacial semelhante a determinado gene conhecido pode revelar uma relação potencial desconhecida, que envolve esses genes em processos e vias semelhantes (NG, Lydia et al., 2007). Neste caso, uma alta correlação de expressão com

um gene circadiano poderia indicar uma relação funcional na regulação central do ritmo circadiano ou simplesmente a expressão circadiana do gene.

Ao nosso conhecimento, não existem, até o momento, nenhum estudo publicado com este tipo de abordagem, para identificar genes envolvidos com o ritmo circadiano.

Na primeira etapa de nossa pesquisa, selecionamos genes candidatos à reguladores dos ritmos circadiano de *M. musculus* e *H. sapiens*, através de mineração de dados experimentais de expressão gênica em larga escala distribuídos espacialmente em diferentes porções do cerebelo destas espécies. Para isto, utilizamos a ferramenta *NeuroBlast*, disponível no portal *Allen Brain Atlas* (<http://www.brain-map.org/>), para identificar genes correlacionados com 19 genes circadianos, que consistem em fatores de transcrição e proteínas de fosforilação e ubiquitinação (*Clock*, *Bmal1*, *Bmal2*, *Cry1*, *Cry2*, *Csnk1a*, *Csnk1d*, *Csnk1e*, *Fbxl3*, *Fbxl21*, *Npas2*, *Nr1d1*, *Nr1d2*, *Per1*, *Per2*, *Per3*, *Rora*, *Rorb* e *Rorc*).

O portal *Allen Brain Atlas* é um recurso de bioinformática que reúne pesquisas em níveis genômicos, com expressão espacial de genes e imagens anatômicas no cérebro. Além disso, disponibiliza dados públicos e de fácil acessibilidade com estudos sobre a organização e funcionamento do cérebro de murino e humanos. LEIN, Ed S. et al. (2007) descreve o *Allen Mouse Atlas* como uma abordagem altamente confiável que pode ajudar a entender a genômica funcional do sistema nervoso, contendo dados de expressão de aproximadamente 20.000 genes de todo o cérebro de camundongos. HAWRYLYCZ, Michael J. et al. (2012) descreve o *Allen Human Brain Atlas* como o único recurso que contém dados globais de expressão gênica de todos os genes e estrutura do cérebro de humanos saudáveis e expressão de centenas genes em cérebro de humanos com distúrbios neurológicos, fornecendo informações importantes para compreender o impacto de doenças genéticas e circuitos funcionais em diferentes regiões do cérebro.

O cerebelo é um relógio circadiano periférico envolvido com o ritmo alimentar de camundongos (MENDOZA, Jorge et al., 2010). É descrito também que os genes circadianos centrais *Per1*, *Per2*, *Per3*, *Cry1*, *Cry2*, *Clock*, *Bmal1*, e *Nr1d1* e um gene controlado pelo relógio, são expressos em células granulares e de Purkinje no córtex cerebelar de ratos. Dentre estes, *Per1*, *Per2*, *Per3*, *Bmal1*, *Cry1*, *Nr1d1*, e *Dbp* apresentaram significantes ritmos diurnos de expressão no cerebelo de animais submetidos em condições de claro e escuro,

assim como também apresentaram padrões cíclicos de expressão em animais submetidos a condições de escuro constante (RATH, Martin F.; ROHDE, Kristian; MOLLER, Morten.,2012).

Outro estudo demonstra que a taxa de disparo neuronal *ex vivo* das células *Purkinje*, presentes no cerebelo de camundongos, apresentam oscilação durante aproximadamente três dias (MORDEL, Jérôme et al.2013). Além disso, existem estudos de transcriptoma circadiano para esta região, em camundongos (ZANGH et al.,2014) e humanos (LI et al.,2013), que identificaram 726 genes e 444 genes com oscilação circadiana nesta estrutura, respectivamente.

Por fim, dados de expressão gênica espacial em larga escala nesta estrutura obtidos através HIS para *M. musculus* e *Microarray* para *H. sapiens* são disponibilizados no portal *Allen Brain Atlas*. Assim, é possível utilizar estas informações para validar pesquisas e comparar genes correlacionados nas duas espécies e validar os dados obtidos com estudos em larga escala já publicados.

Desta forma, identificamos milhares de genes que possuem perfis de expressão espacial similar ou oposta em cerebelo de *M. musculus* e *H. sapiens*, com os 19 genes relógio, muitos destes com valores de r próximos de 1 ou -1. A maior frequência relativa de genes que apresentaram oscilação circadiana demonstrada previamente em cerebelo de camundongos e humanos, com base nos estudos de ZHANG et al.,2014 e LI, J. Z. et al.,2013, está distribuída de acordo com nossos dados, nas faixas com valores mais próximos a 1 ou -1 em ambas as espécies (Figuras 12 e 13). Esta observação é comprovada em *M. Musculus*, quando comparamos as médias dos valores *de r* em módulo ($|r|$) destes grupos de genes que apresentam e não apresentam oscilação circadiana em cerebelo com base no estudo de ZHANG et al.,2014, e encontramos diferença estatisticamente significativa entre eles, onde a média para o grupo que apresenta variação circadiana foi maior ($p < 0,0001$). Porém, comparamos as médias dos valores *de r* em módulo ($|r|$) destes grupos de genes, que apresentam e não apresentam oscilação circadiana em cerebelo, com base no estudo de Li. et al. 2013, e não encontramos diferença estatística para a média dos valores de $|r|$ em humanos. Apesar do estudo de Li. et al. 2013 ser de larga escala, com avaliação de aproximadamente 12.000 transcritos, é um quantitativo bem inferior ao total de genes preditos para a espécie e de transcritos avaliados por ZHANG et al.,2014, com um total de 19. 788. Esta limitação pode

diminuir o percentual de genes encontrados em nossos dados quando comparado com este estudo.

Além disto, a comparação de 2% dos genes correlacionados que possuem os maiores e menores valores de r , mostrou significância estatística para *M. musculus*, nos estudos circadianos de transcriptoma cerebelar, transcriptoma em diferentes tecidos, análises de *ChiP-seq* e proteoma (Figura 14). Já nas análises de 2% dos genes com maiores e menores valores de correlação negativa, em nenhum dos estudos circadianos foi encontrado diferença estatísticas entre os grupos. Acreditamos que este resultado pode estar relacionado ao número de genes correspondentes os 2% selecionado, pois nos grupos com maiores e menores valores de correlação positiva, este percentual corresponde a 370 genes, já para nos grupos com maiores e menores valores de correlação negativa, o percentual corresponde a 27 genes. Além disto, na faixa dos valores de r em que estes 2% dos genes se enquadra, está dentro de um intervalo entre -0,477 a -0,623, menor que o intervalo dos 2% dos genes correlacionados que possuem valores de r positivos, neste caso, entre 0,542 e 0,729 (Figura 15).

Na espécie *H. sapiens*, a comparação de 2% dos grupos de genes com maiores e menores valores de r positivos, mostrou significância estatística nos estudos circadianos de transcriptoma em diferentes tecidos, análises de *ChiP-seq*, proteoma e interação proteína-proteína (Figura 16). Nas análises de 2% dos genes com maiores e menores valores de r negativos, foi identificada significância estatística em quase todos os estudos circadianos avaliados, exceto para o estudo com RNAi (Figura 17). Diante disto, acreditamos que genes com valores de correlações altas negativas correlacionadas aos principais genes circadianos, podem também ser candidatos a apresentar envolvimento com estes ritmos. Todos estes resultados indicam que os genes que apresentam maiores valores de $|r|$ tendem a apresentar envolvimento com os ritmos circadianos.

Dentre os 100 genes melhor ranqueados em nossos dados, encontram-se *Per2* e *Cry2*, com o mesmo valor de correlação, *Arntl* e *Rora*, conhecidos genes circadianos, o que mostra consistência na abordagem utilizada. Porém, cinco entre os 10 primeiros genes mais bem ranqueados não apresentaram nenhum tipo de evidência de envolvimento com o ritmo, incluindo os dois primeiros (*Cartpt* e *Gpr149*). De fato, 27% dos genes com maior correlação não apresentaram nenhuma evidência de envolvimento com o ritmo circadiano nos diferentes estudos experimentais em larga escala considerados (transcriptoma, proteoma, *ChiP-seq*, interação proteína-proteína e RNAi).

É possível que esses genes realmente não apresentem associações com ritmo circadiano. Entretanto, existem inconsistências em dados de avaliação em larga escala publicados, sugerindo limitações metodológicas que poderiam explicar esses achados conflituosos. KOJIMA e colaboradores 2014 apresentam uma crítica a estudos de transcriptoma circadiano. Segundo os autores, estudos que utilizam a mesma abordagem metodológica e avaliação no mesmo tecido biológico, apresentam incompatibilidade entre os resultados. Essas inconsistências também abrangem outros estudos com análises em larga escala como RNA-seq, Microarray, proteoma e Nascent-seq. Os autores chamam a atenção para os diversos algoritmos utilizados na detecção rítmica dos genes, como o *JTK_CYCLE*, *CircWave*, *COSOPT*, dentre outros e que cada um deles tem características específicas para detectar a ritmicidade dos genes. Além disto, processos de normalização de dados de expressão gênica podem introduzir vieses na detecção de ritmicidade. Alguns genes podem ter uma baixa amplitude e, dependendo da sensibilidade do método, a variação de expressão pode não ser identificada (KOJIMA, Shihoko; GREEN, Carla B,2014).

Outros genes podem apresentar variação na expressão apenas em nível proteico e os trabalhos selecionados para análise podem não ter identificado esta variação no tecido. Estudos que realizaram uma comparação entre transcriptoma e proteoma no mesmo tecido sugerem que cerca de 20% do genoma tem variação apenas no nível proteico (REDDY, Akhilesh B. et al.2006), indicando mecanismos pós-transcricionais de regulação da expressão gênica.

Uma outra hipótese que deve ser levada em consideração é a possibilidade do gene não apresentar oscilação circadiana em nenhum nível (dados de transcriptoma e proteoma), não ser regulado por fatores de transcrição circadiano (dados de ChIP-seq), mas sua atividade apresentar variação temporal. Exemplos de genes que não apresentam variação circadiana, ao menos não detectada em alguns tecidos, mas que tem atividade circadiana são *Clock* e quinases CK1 ϵ , CK1 δ e AMP. Estas últimas são ativadas circadianamente atuam fosforilando as proteínas PER e CRY (LOWREY e TAKAHASHI, 2004; TAKAHASHI *et al.*, 2008; MOHAWK, Jennifer A.; GREEN, Carla B.; TAKAHASHI, Joseph S.,2012;). Além destas, outras quinases como a PKA, CaMK II e a MAPK estão envolvidas na regulação da ativação da proteína CLOCK (WEBER, Frank et al., 2006).

Incluimos dados de interação proteína-proteína com componentes do relógio, coletados a partir de portal *Protein Network Analysis* (PINA), mas apenas 8 genes apresentaram associação: *Gcfc1*, *Pcyox1*, *Parp1* e *Rpl27a*, e os genes circadianos *Per2*, *Cry2*, *Bmal1* e *Rora*.

As informações coletadas de ChiP-seq para proteínas circadianas foram obtidas a partir de diferentes estudos (KOIKE et al., 2013; FANG et al., 2014). No entanto, os trabalhos publicados até o momento acessaram esta interação DNA-proteína apenas em um único tecido (fígado), o que restringe as possíveis variações epigenéticas que podem ocorrer em diferentes contextos.

Os dados obtidos de um único estudo em larga escala de *screening* com RNAi para genes circadiano, publicado até o momento, relacionou apenas 3 genes na nossa lista de 100 genes com maior correlação, que também foram associados com uma interferência no ritmo circadiano (*Cry2*, *Abll* e *Homer3*). Entretanto, os genes incluídos na lista para esta análise estão entre aqueles previamente selecionados por ZHANG et al., 2009, com critérios bastante rigorosos para interferência nos fenótipos circadianos. Além disto, o estudo foi realizado em um modelo *in vitro*, com células contendo os genes repórter *Bmal1-dLuc* e *Per2-dLuc*, derivadas da linhagem celular U2OS, o que certamente restringe o número de genes potencialmente envolvidos com o ritmo *in vivo* no cérebro. Esta limitação também pode estar associada a diferença estatística não encontrada, entres os grupos de 2% de genes com maiores e menores valores de correlação positiva e negativa nas espécies estudadas.

Considerando estas limitações, a abordagem utilizada de detecção de associação com o ritmo circadiano através de correlação da expressão de transcritos, de acordo com a distribuição espacial no cérebro, tem potencial para identificação de novos genes candidatos. A análise do conjunto de evidências prévias sobre o envolvimento com ritmos circadianos sugere que o método pode detectar de uma maneira diferente genes não caracterizados previamente como circadianos. Isto pode ser feito para todo o cérebro ou para estruturas específicas ainda não exploradas, em camundongos e cérebro humano, sobrepujando limitações experimentais de outros métodos. A análise conjunta destas espécies pode ainda permitir a identificação de genes homólogos apresentando relações funcionais conservadas com o ritmo circadiano.

Identificamos, em ambas as espécies, diferentes padrões de correlações altas, médias e baixas, entre os 19 genes circadianos estudados. Isto pode ser explicado, a princípio, pelas diferentes interações funcionais que estas moléculas estabelecem entre si, em nível transcricional e pós-transcricional. *Per2* e *Cry2* apresentam correlação recíproca, com valor de r de 0,603, uma das maiores correlações encontradas em cerebelo de camundongos e com uma alta correlação de 0,855 em cerebelo de humanos. De fato, com o aumento dos níveis das proteínas PERs e CRYs, são formados heterodímeros, que irão inibir a transcrição de outros genes circadianos, a partir da inibição do seu complexo ativador, o heterodímero CLOCK:BMAL1, constituindo desta forma, uma alça de retroalimentação negativa nesse sistema (TAKAHASHI, J. S. 2015; PARTCH, Carrie L.; GREEN, Carla B.; TAKAHASHI, Joseph S.2014).

Identificamos também que *Rora* e *Bmal1* estabelecem uma das mais altas correlações em camundongos, com valor de r de 0,628. Esta relação pode ser explicada através da interação que estes genes realizam entre si no sistema molecular de temporização circadiana de mamíferos. CLOCK:BMAL1 formam um heterodímero que atua na alça de retroalimentação positiva desse sistema, ativando a transcrição de diversos genes, inclusive os *Rors*. Por outro lado, as proteínas dos genes *Rors*, ativam a transcrição do gene *Bmal1*, constituindo uma segunda alça de retroalimentação positiva nesse sistema (TAKAHASHI, J. S. 2015; PARTCH, Carrie L.; GREEN, Carla B.; TAKAHASHI, Joseph S.2014).

Evidenciamos uma correlação alta entre os genes *Nr1d2* e o gene *Bmal1*, r de 0,732, para as análises das correlações em cerebelo de *H.sapiens* (Tabela 02). O heterodímero CLOCK:BMAL1 ativa a transcrição dos genes *Nr1d1* e *Nr1d2*, neste caso, maiores níveis de BMAL1 implicam em maiores níveis de *Nr1d2*. Por outro lado, as proteínas NR1D1 e NR1D2 atuam na alça de retroalimentação negativa, competindo com as proteínas RORs pelo mesmo sítio de ligação do promotor de *Bmal1*, reprimindo desta forma, seu processo de transcrição, o que poderia resultar em uma correlação negativa (TAKAHASHI, J. S. 2015; PARTCH, Carrie L.; GREEN, Carla B.; TAKAHASHI, Joseph S.2014). De fato, também evidenciamos uma correlação baixa negativa entre o gene *Bmal2* e o *Nr1d1*, com valor de r de -0,319.

O gene *Npas2* também apresentou correlações com valores de r baixos com genes circadianos estudados. No sistema molecular de temporização circadiana de mamífero este

gene atua como análogo do gene *Clock*, na falta deste último, o gene *Npas2* assume sua função neste sistema (REICK, Martin et al., 2001).

Algumas proteínas participam dos mecanismos pós-transicionais que atuam no sistema de temporização circadiano de mamíferos, através dos processos de fosforilação e ubiquitinação de outras proteínas. As quinases (CSNK1A, CSNK1D e CSNK1E) atuam fosforilando as proteínas PER e CRY. Encontramos em nossos resultados correlações altas e médias entre os genes dessas proteínas em cerebelo de *H. sapiens*. Por exemplo, o gene *Csnk1a*, apresentou um valor de correlação de 0,761 e 0,729, com os genes *Cry1* e *Cry2*, respectivamente. O gene *Csnk1d* apresentou correlações com valores de -0,809, -0,652 e -0,781 com os genes *Per1*, *Per2* e *Per3*, respectivamente. O gene *Csnk1e* apresentou correlações com valores de 0,632, 0,718, 0,658 e 0,636 com os genes *Cry2*, *Per1*, *Per2* e *Per3*, respectivamente.

As proteínas FBXL3 e FBL21 atuam degradando PER e CRY, através do processo de poliubiquitinação, juntamente com o complexo ubiquitinaligase E3 SCF, a enzima β -TrCP e o complexo proteossomo 26S (LOWREY e TAKAHASHI, 2004; TAKAHASHI *et al.*, 2008). Em nossos resultados, evidenciamos correlação média entre o gene *Fbxl3* e o gene *Per1*, com valor de r de 0,416 para a espécie *M. Musculus* e de -0,412 para *H. sapiens*. Estas correlações, positiva e negativa, tem relação com a interação que estes genes estabelecem no sistema molecular e possivelmente com o ensaios utilizados em cada caso. Os genes *Pers* e *Pblx3* são regulados pelo mesmo fator de transcrição CLOCK:BMAL, que resulta em aumento ou diminuição simultâneos da transcrição dos genes *Pers* e *Pblx3*. Acreditamos que essa possibilidade justifica a correlação positiva encontrada entre esses genes. (SIEPKA, Sandra M. et al., 2007; MOHAWK, Jennifer A.; GREEN, Carla B.; TAKAHASHI, Joseph S., 2012; TAKAHASHI, J. S. 2015). Além disso, como são genes que possuem expressão oscilatória e as correlações são geradas em duas espécies com pareamentos de ensaios diferentes, é possível que a expressão desses genes tenha sido avaliada em fases distintas do ritmo e, neste caso, ter resultado em uma correlação negativa. O gene *Fbl21* apresentou correlações com valores de r baixos com genes circadianos estudados nos dados de ambas as espécies.

De qualquer forma, os dados de correlações recíprocas entre os genes circadianos mostram que o simples fato de ser um gene circadiano não implica correlação alta e que isto depende das interações funcionais estabelecidas. Isto pode indicar uma baixa especificidade do método em incluir genes candidatos a partir do conjunto de dados. Por outro lado, as altas

correlações identificadas são compatíveis com as interações biológicas conhecidas para estes genes, o que sugere que o método possui alta sensibilidade na identificação de relações funcionais.

7.1 GENES CIRCADIANOS CANDIDATOS IDENTIFICADOS EM CEREBELO

Cartpt, o gene ranqueado na primeira posição, apesar de não ter apresentado nenhuma evidência em estudos em larga escala, foi associado com o ritmo alimentar de mamíferos. VICENTIC, A. et al., (2004) dosaram os níveis diurnos de peptídeos de *Cartpt* através da radioimunoensaio em sangue de ratos em jejum por 24 horas, identificando 50 pg/ml desses peptídeos nas horas iniciais da manhã e 150 pg/ml no início da noite. Além disso, avaliaram a expressão da proteína através de *Western blotting*, na qual observaram baixos níveis de proteína no início da manhã. Os peptídeos de *Cartpt* exibem, assim, um ritmo diurno em sangue de ratos. VICENTIC, Aleksandra et al.,(2005) demonstram também o ritmo diurno dos peptídeos de *Cartpt* em regiões do cérebro e que os mesmos são afetados pelo ritmo alimentar, encontrando que nos núcleos accumbens, hipotálamo e amígdala de camundongos submetidos a jejum de 24 horas os níveis dos peptídeos são baixos e pela manhã e aumentados no início da noite. Além disso, demonstraram que a expressão do RNAm de *Cartpt* apresenta variação circadiana em núcleos accumbens de animais submetidos a jejum. Em suma, os resultados encontrados demonstram que níveis dos peptídeos de expressão do gene apresentam variação diurna em algumas regiões do cérebro, e que essa variação é alterada pelo ritmo alimentar.

Entretanto, STÜTZ, Adrian M. et al., (2007) realizou um estudo para avaliar a expressão circadiana de genes envolvidos com o ritmo alimentar em camundongos, inclusive *Cartpt*, demonstrando que este gene não exibiu expressão circadiana nos hipotálamos de camundongos, e não foi envolvido com ritmo circadiano alimentar. Considerando estas evidências, a expressão do RNAm de *Cartpt* poderia não ser rítmica em cerebelo, mas a proteína apresentar oscilação circadiana, o que explicaria a maior correlação deste gene com o ritmo circadiano.

Parp1 (poly ADP-ribosyl) apresenta expressão circadiana, já descrita no fígado de camundongos e também é regulado pelo ritmo alimentar. Camundongos *knockout* para este gene tiveram arrastamento de fase do ritmo alimentar. Além disso, o *Parp1* interage com o

heterodímero CLOCK: BMAL1 modulando a atividade desses fatores de transcrição (ASHER, Gad et al.,2010).

Os genes *Cnr1*, *Inpp4a*, *Prdm10*, *Gnao1* e *Baiap2* também estão entre os 10 mais bem ranqueados e apresentaram outras evidências nos estudos circadianos avaliados (transcriptoma, ChiP-seq).

Grid2 está situado na décima posição do nosso ranque e não apresentou nenhuma evidência nos estudos circadianos avaliados. No entanto, é comprovada a relação deste gene está com o ritmo alimentar em camundongos. A atividade antecipatória alimentar é um ritmo observado em camundongos, após submetê-los em condições de restrição de alimentos. MENDOZA e colaboradores, 2010, a fim de comprovar a relação do gene *Grid2* coma atividade antecipatória alimentar, realizou experimentos com camundongos mutantes heterozigotos e homozigotos para este gene e comparou essa atividade, o movimento do animais na rodana e a temperatura corporal, nesses dois fenótipos com camundongos normais, durante 20 dias de restrição alimentar, disponibilizando 60% de alimento durante todo dia às 12:00 horas. Eles identificaram através de actogramas que os animais mutantes heterozigotos apresentam diminuição dessa atividade antecipatória, no movimento na rodana e temperatura corporal e os homozigotos quase não apresentaram esses dois tipos de atividade e temperaturas corporais mais baixas. Além disso, ter certeza que essa diminuição nesses parâmetros não está sendo mascarada pela luz e sim é, ocasionada por um mecanismo endógeno, eles submeteram os animais mutantes homozigotos nessas condições de restrição alimentar comparando com os normais e comprovaram os resultados descritos anteriormente. Além disso, eles avaliam a expressão dos genes circadianos *Per1*, *Per2*, *Nr1d1* e *Dbp*, em cerebelo de animais normais e nesses mutantes, nestas condições de restrição e identificam uma diminuição nos níveis de expressão desses genes.

O gene *Homer3* não possui evidências de seu envolvimento com os ritmos circadianos, mas um membro da mesma família (*Homer1a*), é descrito como um gene relógio envolvido com o ritmo de sono (MARET, Stéphanie et al.,2007). *Npas3* também não apresenta evidências de participação no ritmo circadiano, mas é homólogo a *Npas2*, que funciona como um gene análogo ao gene *Clock*, atuando como um agente compensatório dos ritmos em NSQ na ausência deste gene (REICK, Martin et al,2001; DEBRUYNE, Jason P.,2008).

Desta forma, através banco de dados *Allen Brain Atlas* foi possível em nossa pesquisa identificar novos genes potencialmente importantes para o controle do ritmo circadiano, utilizando análises de correlação de expressão espacial e mineração de dados com estudos de larga escala circadianos. No entanto, a similaridade de perfil de expressão espacial com os genes circadianos, que já tem expressão circadiana comprovada experimentalmente e envolvimento com ritmos, por si só, não nos garante que estes genes correlacionados também apresentem as mesmas funções. Sendo assim, outros estudos de expressão gênica e ensaios funcionais são necessários para confirmar o envolvimento dos genes candidatos com o sistema de temporização circadiana de mamíferos.

8 CONCLUSÕES

Foram identificados até o momento um conjunto de genes candidatos a fazer parte do sistema de temporização circadiana em cerebelo humano e murino, alguns com comprovação experimental já publicada;

A maioria dos genes identificados com alta correlação apresenta variação circadiana identificada previamente em estudos de análise de transcriptoma;

O gene *Cartpt* foi o gene com maior valor de correlação identificado, considerando os dados das duas espécies, e constitui-se como um forte candidato a controlar o ritmo alimentar em cerebelo de mamíferos;

Os genes *Gpr149*, *Ptprr*, *Nipal4* estão entre os 10 mais bem ranqueados e não apresentam nenhuma evidência prévia de envolvimento com o ritmo constituindo-se como novos candidatos circadianos em cerebelo de mamíferos;

Este é o primeiro trabalho que utiliza a estratégia de correlação de expressão espacial e o banco de dados *Allen Brain Atlas* para buscar genes com expressão circadiana e envolvimento com os ritmos circadianos;

A estratégia utilizada tem potencial para identificar genes importantes para o ritmo circadiano em estruturas específicas do cérebro humano e de camundongos, sobrepujando limitações metodológicas consideradas em outros estudos;

Análises de expressão gênica e ensaios funcionais são necessários para validar o envolvimento destes genes candidatos.

REFERÊNCIAS

ABE, M. et al. Circadian rhythms in isolated brain regions. **The Journal of neuroscience : the official journal of the Society for Neuroscience**, v. 22, n. 1, p. 350–356, 2002.

ABRAHAMSON, Eric E.; MOORE, Robert Y. Suprachiasmatic nucleus in the mouse: retinal innervation, intrinsic organization and efferent projections. **Brain research**, v. 916, n. 1, p. 172-191, 2001.

AKASHI, Makoto; TAKUMI, Toru. The orphan nuclear receptor ROR α regulates circadian transcription of the mammalian core-clock *Bmal1*. **Nature structural & molecular biology**, v. 12, n. 5, p. 441-448, 2005.

ALBRECHT, Urs; EICHELE, Gregor. The mammalian circadian clock. **Current opinion in genetics & development**, v. 13, n. 3, p. 271-277, 2003.

ALMON, Richard R. et al. Relationships between circadian rhythms and modulation of gene expression by glucocorticoids in skeletal muscle. **American Journal of Physiology-Regulatory, Integrative and Comparative Physiology**, v. 295, n. 4, p. R1031-R1047, 2008

ANAFI, Ron C. et al. Machine learning helps identify CHRONO as a circadian clock component. **PLoS Biol**, v. 12, n. 4, p. e1001840, 2014.

ASHER, Gad et al. Poly (ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. **Cell**, v. 142, n. 6, p. 943-953, 2010.

BELL-PEDERSEN, Deborah et al. Circadian rhythms from multiple oscillators: lessons from diverse organisms. **Nature Reviews Genetics**, v. 6, n. 7, p. 544-556, 2005.

BHARGAVA, Anuprabha; HERZEL, Hanspeter; ANANTHASUBRAMANIAM, Bharath. Mining for novel candidate clock genes in the circadian regulatory network. **BMC systems biology**, v. 9, n. 1, p. 1, 2015.

BOTBOL, Michel et al. Biological and psychological rhythms: An integrative approach to rhythm disturbances in autistic disorder. **Journal of Physiology-Paris**, v. 107, n. 4, p. 298-309, 2013.

BROOKS, Elisabeth; CANAL, Maria M. Development of circadian rhythms: Role of postnatal light environment. **Neuroscience & Biobehavioral Reviews**, v. 37, n. 4, p. 551-560, 2013.

BUHR, Ethan D.; TAKAHASHI, Joseph S. Molecular components of the mammalian circadian clock. In: **Circadian clocks**. Springer Berlin Heidelberg, 2013. p. 3-27.

BUNGER, Maureen K. et al. Mop3 is an essential component of the master circadian pacemaker in mammals. **Cell**, v. 103, n. 7, p. 1009-1017, 2000.

CHIANG, C. K. et al. The Proteomic Landscape of the Suprachiasmatic Nucleus Clock Reveals Large-Scale Coordination of Key Biological Processes. **PLoS Genetics**, v. 10, n. 10, 2014.

DAMIOLA, Francesca et al. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus.

Genes & development, v. 14, n. 23, p. 2950-2961, 2000.

DARDENTE, Hugues et al. Implication of the F-Box Protein FBXL21 in circadian pacemaker function in mammals. **PLoS One**, v. 3, n. 10, p. e3530, 2008.

DIBNER, Charna; SCHIBLER, Ueli; ALBRECHT, Urs. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. **Annual review of physiology**, v. 72, p. 517-549, 2010.

DUDLEY, Carol A. et al. Altered patterns of sleep and behavioral adaptability in NPAS2-deficient mice. **Science**, v. 301, n. 5631, p. 379-383, 2003.

DUNLAP, Jay C. Molecular bases for circadian clocks. **Cell**, v. 96, n. 2, p. 271-290, 1999.

FANG, B. et al. Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. **Cell**, v. 159, n. 5, p. 1140–1152, 2014.

FEILLET, Céline A. et al. Forebrain oscillators ticking with different clock hands. **Molecular and Cellular Neuroscience**, v. 37, n. 2, p. 209-221, 2008.

FILICHKIN, Sergei A. et al. Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. **PloS one**, v. 6, n. 6, p. e16907, 2011.

GUILDING, Clare, and HUGH D. Piggins. "Challenging the omnipotence of the suprachiasmatic timekeeper: are circadian oscillators present throughout the mammalian brain?" **European Journal of Neuroscience** 25.11 (2007): 3195-3216.

GRANADOS- FUENTES, Daniel et al. Olfactory bulb neurons express functional, entrainable circadian rhythms. **European Journal of Neuroscience**, v. 19, n. 4, p. 898-906, 2004.

GRIFFIN, Edmund A.; STAKNIS, David; WEITZ, Charles J. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. **Science**, v. 286, n. 5440, p. 768-771, 1999.

HAND, David J.; MANNILA, Heikki; SMYTH, Padhraic. **Principles of data mining**. MIT press, 2001.

HASTINGS, Michael H.; REDDY, Akhilesh B.; MAYWOOD, Elizabeth S. A clockwork web: circadian timing in brain and periphery, in health and disease. **Nature Reviews Neuroscience**, v. 4, n. 8, p. 649-661, 2003.

HASTINGS, M. H. et al. Expression of clock gene products in the suprachiasmatic nucleus in relation to circadian behaviour. In: **Novartis Found Symp.** 2004. p. 203-217.

HAWRYLYCZ, Michael J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. **Nature**, v. 489, n. 7416, p. 391-399, 2012.

HERZOG, Erik D. Neurons and networks in daily rhythms. **Nature Reviews Neuroscience**, v. 8, n. 10, p. 790-802, 2007.

HONMA, S. et al. Dec1 and Dec2 are regulators of the mammalian molecular clock. **Nature**, v. 419, n. 6909, p. 841–844, 2002.

- HOGENESCH, John B.; UEDA, Hiroki R. Understanding systems-level properties: timely stories from the study of clocks. **Nature Reviews Genetics**, v. 12, n. 6, p. 407-416, 2011.
- HUGHES, M. E. et al. Harmonics of Circadian Gene Transcription in Mammals. v. 5, n. 4, 2009.
- JONES, Allan R.; OVERLY, Caroline C.; SUNKIN, Susan M. The Allen brain atlas: 5 years and beyond. **Nature Reviews Neuroscience**, v. 10, n. 11, p. 821-828, 2009.
- KO, Caroline H.; TAKAHASHI, Joseph S. Molecular components of the mammalian circadian clock. **Human molecular genetics**, v. 15, n. suppl 2, p. R271-R277, 2006.
- KOIKE, N. et al. NIH Public Access. v. 338, n. 6105, p. 349–354, 2013.
- KOJIMA, Shihoko; SHINGLE, Danielle L.; GREEN, Carla B. Post-transcriptional control of circadian rhythms. **Journal of cell science**, v. 124, n. 3, p. 311-320, 2011.
- KOJIMA, Shihoko; GREEN, Carla B. Circadian genomics reveal a role for post-transcriptional regulation in mammals. **Biochemistry**, v. 54, n. 2, p. 124-133, 2014.
- KUAN, Leonard et al. Neuroinformatics of the allen mouse brain connectivity atlas. **Methods**, v. 73, p. 4-17, 2015.
- KUELLER, Rikard. The influence of light on circarrhythms in humans. **Journal of physiological anthropology and applied human science**, v. 21, n. 2, p. 87-91, 2002.
- LAING, Emma E. et al. Exploiting human and mouse transcriptomic data: Identification of circadian genes and pathways influencing health. **BioEssays**, v. 37, n. 5, p. 544-556, 2015.
- LEHMANN, Robert et al. Assembly of a comprehensive regulatory network for the mammalian circadian clock: a bioinformatics approach. **PloS one**, v. 10, n. 5, p. e0126283, 2015.4
- LEIN, Ed S. et al. Genome-wide atlas of gene expression in the adult mouse brain. **Nature**, v. 445, n. 7124, p. 168-176, 2007.
- LEMMER, Björn. Discoveries of rhythms in human biological functions: a historical review. **Chronobiology international**, v. 26, n. 6, p. 1019-1068, 2009.
- LOPES, Robson da Silva et al. Application of bioinformatics in chronobiology research. **The Scientific World Journal**, v. 2013, 2013.
- LI, J. Z. et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. 2013.
- LOWREY, Phillip L.; TAKAHASHI, Joseph S. Mammalian circadian biology: elucidating genome-wide levels of temporal organization. **Annual review of genomics and human genetics**, v. 5, p. 407, 2004.
- LOWREY, Phillip L.; TAKAHASHI, Joseph S. Genetics of circadian rhythms in Mammalian model organisms. **Advances in genetics**, v. 74, p. 175, 2011.
- MAIER, Bert et al. A large-scale functional RNAi screen reveals a role for CK2 in the mammalian circadian clock. **Genes & Development**, v. 23, n. 6, p. 708-718, 2009.

- MALER, Thomas et al. The *Drosophila* per gene homologs are expressed in mammalian suprachiasmatic nucleus and heart as well as in molluscan eyes. **Biochemical and biophysical research communications**, v. 184, n. 2, p. 1082-1087, 1992.
- MARET, Stéphanie et al. Homer1a is a core brain molecular correlate of sleep loss. **Proceedings of the National Academy of Sciences**, v. 104, n. 50, p. 20090-20095, 2007.
- MARQUES, Nelson; BARRETO, Luiz Menna. **Cronobiologia: Princípios e Aplicações**. 3ed. Edusp. São Paulo, 2003. ISBN: 85-314-0400-2.
- MAUVOISIN, D. et al. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 1, p. 167-72, 2014.
- MENDOZA, Jorge; CHALLET, Etienne. Brain clocks: from the suprachiasmatic nuclei to a cerebral network. **The Neuroscientist**, v. 15, n. 5, p. 477-488, 2009.
- MENDOZA, Jorge et al. The cerebellum harbors a circadian oscillator involved in food anticipation. **The Journal of Neuroscience**, v. 30, n. 5, p. 1894-1904, 2010.
- MOHAWK, Jennifer A.; GREEN, Carla B.; TAKAHASHI, Joseph S. Central and peripheral circadian clocks in mammals. **Annual review of neuroscience**, v. 35, p. 445, 2012.
- MORDEL, Jérôme et al. The output signal of Purkinje cells of the cerebellum and circadian rhythmicity. **PloS one**, v. 8, n. 3, p. e58457, 2013.
- NAMIHIRA, Masakazu et al. Daily variation and light responsiveness of mammalian clock gene, Clock and BMAL1, transcripts in the pineal body and different areas of brain in rats. **Neuroscience letters**, v. 267, n. 1, p. 69-72, 1999.
- NAYAK, Renuka R. et al. Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. **Genome research**, v. 19, n. 11, p. 1953-1962, 2009.
- NG, Lydia et al. NeuroBlast: a 3D spatial homology search tool for gene expression. **BMC Neuroscience**, v. 8, n. 2, p. 1, 2007.
- NG, Lydia et al. Neuroinformatics for genome-wide 3-D gene expression mapping in the mouse brain. **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, v. 4, n. 3, p. 382-393, 2007.
- NG, Lydia et al. An anatomic gene expression atlas of the adult mouse brain. **Nature neuroscience**, v. 12, n. 3, p. 356-362, 2009.
- PARTCH, Carrie L.; GREEN, Carla B.; TAKAHASHI, Joseph S. Molecular architecture of the mammalian circadian clock. **Trends in cell biology**, v. 24, n. 2, p. 90-99, 2014.
- PEREIRA, Danyella Silva; TUFIK, Sergio; PEDRAZZOLI, Mario. Moléculas que marcam o tempo: implicações para os fenótipos circadianos Timekeeping molecules: implications for circadian phenotypes. **Rev Bras Psiquiatr**, v. 31, n. 1, p. 63-71, 2009.
- PREITNER, N. et al. Orphan nuclear receptors, molecular clockwork, and the entrainment of peripheral oscillators. **Novartis Found Symp**, v. 253, p. 89-99; discussion 99-109, 2003.

- PREITNER, N. et al. The orphan nuclear receptor REV-ERB α controls circadian transcription within the positive limb of the mammalian circadian oscillator. **Cell**, v. 110, n. 2, p. 251-60, Jul 26. 2002.
- RATH, Martin F.; ROHDE, Kristian; MØLLER, Morten. Circadian oscillations of molecular clock components in the cerebellar cortex of the rat. **Chronobiology international**, v. 29, n. 10, p. 1289-1299, 2012.
- REDDY, A. B. et al. Circadian orchestration of the hepatic proteome. **Curr Biol**, v. 16, n. 11, p. 1107-15, Jun 6. 2006.
- REICK, Martin et al. NPAS2: an analog of clock operative in the mammalian forebrain. **Science**, v. 293, n. 5529, p. 506-509, 2001.
- ROBLES, M. S.; COX, J.; MANN, M. In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. **PLoS Genetics**, v. 10, n. 1, 2014.
- ROSSNER, Moritz J. et al. Disturbed clockwork resetting in Sharp-1 and Sharp-2 single and double mutant mice. **PLoS One**, v. 3, n. 7, p. e2762, 2008.
- SATO, T. K. et al. A functional genomics strategy reveals rora as a component of the mammalian circadian clock. **Neuron**, v. 43, n. 4, p. 527–537, 2004.
- SHEARMAN, Lauren P. et al. Interacting molecular loops in the mammalian circadian clock. **Science**, v. 288, n. 5468, p. 1013-1019, 2000.
- SHEN, Elaine H.; OVERLY, Caroline C.; JONES, Allan R. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. **Trends in neurosciences**, v. 35, n. 12, p. 711-714, 2012.
- SIEPKA, Sandra M. et al. Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. **Cell**, v. 129, n. 5, p. 1011-1023, 2007.
- SUKUMARAN, Siddharth et al. Circadian variations in gene expression in rat abdominal adipose tissue and relationship to physiology. **Physiological genomics**, v. 42, n. 2, p. 141-152, 2010.
- SUNKIN, Susan M. et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. **Nucleic acids research**, p. gks1042, 2012.
- SZKLARCZYK, Damian et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. **Nucleic acids research**, p. gku1003, 2014.
- TAKUMI, T. et al. A new mammalian period gene predominantly expressed in the suprachiasmatic nucleus. **Genes to Cells**, v. 3, n. 3, p. 167–176, 1998.
- TAKAHASHI, Joseph S. et al. The genetics of mammalian circadian order and disorder: implications for physiology and disease. **Nature Reviews Genetics**, v. 9, n. 10, p. 764-775, 2008.
- TAKAHASHI, J. S. Molecular components of the circadian clock in mammals. **Diabetes, Obesity and Metabolism**, v. 17, n. S1, p. 6-11, 2015.

- TEI, Hajime et al. Circadian oscillation of a mammalian homologue of the *Drosophila* period gene. **Nature**, v. 389, n. 6650, p. 512-516, 1997.
- TRIQUENEAUX, Gérard et al. The orphan receptor Rev-erb α gene is a target of the circadian clock pacemaker. **Journal of molecular endocrinology**, v. 33, n. 3, p. 585-608, 2004.
- TOSINI, Gianluca et al. The circadian clock system in the mammalian retina. **Bioessays**, v. 30, n. 7, p. 624-633, 2008.
- WEBER, Frank et al. Second messenger and Ras/MAPK signalling pathways regulate CLOCK/CYCLE- dependent transcription. **Journal of neurochemistry**, v. 98, n. 1, p. 248-257, 2006.
- WELSH, David K.; TAKAHASHI, Joseph S.; KAY, Steve A. Suprachiasmatic nucleus: cell autonomy and network properties. **Annual review of physiology**, v. 72, p. 551, 2010.
- VAN DER HORST, G. T. et al. Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. **Nature**, v. 398, n. 6728, p. 627–30, 1999.
- VIDENOVIC, Aleksandar et al. 'The clocks that time us'[mdash] circadian rhythms in neurodegenerative disorders. **Nature Reviews Neurology**, v. 10, n. 12, p. 683-693, 2014.
- VITATERNA, Martha Hotz et al. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. **Science (New York, NY)**, v. 264, n. 5159, p. 719, 1994.
- YAMAZAKI, Shin et al. Resetting central and peripheral circadian oscillators in transgenic rats. **Science**, v. 288, n. 5466, p. 682-685, 2000.
- ZHANG, E. E. et al. A Genome-wide RNAi Screen for Modifiers of the Circadian Clock in Human Cells. **Cell**, v. 139, n. 1, p. 199–210, 2009.
- ZHANG, R. et al. A circadian gene expression atlas in mammals : Implications for biology and medicine. n. 642, 2014.