



Published in final edited form as:

*J Pineal Res.* 2020 October ; 69(3): e12673. doi:10.1111/jpi.12673.

## Resource: A Multi-species Multi-timepoint Transcriptome Database and Webpage for the Pineal Gland and Retina

**Eric Chang**<sup>1,a,b</sup>, **Cong Fu**<sup>2,3,4,5,a</sup>, **Steven L. Coon**<sup>2,6,a</sup>, **Shahar Alon**<sup>7,c</sup>, **Marjan Bozinoski**<sup>8</sup>, **Matthew Breymaier**<sup>9</sup>, **Diego M. Bustos**<sup>2,d</sup>, **Samuel J. Clokie**<sup>2,e</sup>, **Yoav Gothilf**<sup>7</sup>, **Caroline Esnault**<sup>1,b</sup>, **P. Michael Iuvone**<sup>10</sup>, **Christopher E. Mason**<sup>8</sup>, **Margaret J. Ochocinska**<sup>2,f</sup>, **Adi Tovin**<sup>7,g</sup>, **Charles Wang**<sup>11</sup>, **Pinxian Xu**<sup>12</sup>, **Jinhan Zhu**<sup>13,14</sup>, **Ryan Dale**<sup>1,a,b</sup>, **David C. Klein**<sup>2,15,h</sup>

<sup>1</sup>Bioinformatics and Scientific Programming Core, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892

<sup>2</sup>Section on Neuroendocrinology, Program in Developmental Endocrinology and Genetics, *Eunice Shriver Kennedy* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892

<sup>3</sup>Key Laboratory of Organ Regeneration & Transplantation of the Ministry of Education, The First Hospital of Jilin University, Changchun, 130061, PR China

<sup>4</sup>Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun 130061, PR China

<sup>5</sup>National-Local Joint Engineering Laboratory of Animal Models for Human Diseases, Changchun, 130061, PR China

<sup>6</sup>Molecular Genomics Core, Office of the Scientific Director, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda MD 20892

<sup>7</sup>Department of Neurobiology, The George S. Wise Faculty of Life Sciences, and Sagol School of Neuroscience, Tel-Aviv University, Tel Aviv 69978, Israel

<sup>8</sup>Department of Physiology and Biophysics and the Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY 10021

<sup>9</sup>Computer Support Services Core, *Eunice Shriver Kennedy* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892

---

<sup>h</sup>Corresponding author: National Institutes of Health, Bethesda, Maryland 20892, kleind@mail.nih.gov, 301-496-6915 (off), 301-480-5229 (fax).

<sup>a</sup>These authors contributed equally to this manuscript

<sup>b</sup>These authors were responsible for bioinformatics

<sup>c</sup>Current affiliation: The Alexander Kofkin Faculty of Engineering, Bar-Ilan University, Ramat-Gan, 52900 Israel

<sup>d</sup>Current affiliation: Instituto de Histología y Embriología de Mendoza, Consejo Nacional de Investigaciones Científicas y Técnicas, Mendoza, Argentina

<sup>e</sup>Current affiliation: West Midlands Regional Genetics Laboratories, Birmingham, Women's and Children's NHS Foundation Trust, Birmingham, UK

<sup>f</sup>Current affiliation: National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892

<sup>g</sup>Current affiliation: The Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, 52900 Israel

**Conflicts of interest:** No conflicts of interest related to this manuscript exist.

<sup>10</sup>Departments of Ophthalmology and Pharmacology & Chemical Biology, Emory University School of Medicine, Atlanta, GA 30322

<sup>11</sup>Center for Genomics, School of Medicine, Loma Linda University, Loma Linda, CA 92350

<sup>12</sup>Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine Icahn Medical Institute, New York, NY 10029

<sup>13</sup>United States Food and Drug Administration's National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079

<sup>14</sup>Deptment of Physiology, School of Basic Medical Sciences, Anhui Medical University, Hefei, 230032, China

<sup>15</sup>Office of the Scientific Director, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda MD 20892

## Abstract

The website and database <https://snengs.nichd.nih.gov> provides RNA sequencing data from multi-species analysis of the pineal glands from zebrafish (*Danio rerio*), chicken (White Leghorn), rat (Sprague Dawley), mouse (129sv), rhesus macaque (*Macaca mulatta*) and human (*Homo sapiens*); in most cases retinal data are also included along with results of the analysis of a mixture of RNA from tissues. Studies cover day and night conditions; in addition, a time series over multiple hours, a developmental time series and pharmacological experiments on rats are included. The data have been uniformly re-processed using the latest methods and assemblies to allow for comparisons between experiments and to reduce processing differences. The website presents search functionality, graphical representations, Excel tables, and track hubs of all data for detailed visualization in the UCSC Genome Browser. As more data are collected from investigators and improved genomes become available in the future, the website will be updated. This database is in the public domain and elements can be reproduced by citing the URL or this report. This effort makes the results of 21<sup>st</sup> century transcriptome profiling widely available in a user-friendly format that is expected to broadly influence pineal research.

## Introduction

The pineal transcriptome has been studied for over 30 years, starting with Northern blot detection of single transcripts encoding proteins involved in melatonin synthesis, including those encoding Tph1 and Asmt (Hiomt) <sup>1-4</sup>. Since then, pineal transcriptomics has spanned the development of transcriptomic assays including cDNA-based hybridization technology, qRT-PCR and RNA-Seq <sup>5-15</sup>.

High-throughput sequencing offers many advantages for assaying the transcriptome, but the field of bioinformatics moves quickly with tools, algorithms, and genome assemblies changing from year to year. As a result, data from earlier studies cannot be meaningfully compared with data from later studies that used different methods without completely re-processing all data uniformly using updated methods, assemblies, and annotations. This has been recognized for example in the recount2 project <sup>16</sup>, which has re-processed tens of thousands of human RNA-Seq samples from public repositories uniformly.

Furthermore, for detailed study of particular loci it is critical to visualize expression alongside genomic data from other studies. Genome browsers such as the UCSC Genome Browser <sup>17</sup> allow just this. In particular, this browser supports track hubs that allow for the configuration, coloration and organization of collections of many tracks using a web interface <sup>18</sup>. This allows researchers to generate highly customized views tailored to their research interest, viewing pineal gland data from this study directly alongside a wealth of publicly available data prepared and made available by the UCSC team.

Cross-species transcriptome reports appear in the literature, focused on two-species comparisons, e.g., mouse versus human <sup>19</sup> and zebrafish versus human <sup>20</sup>. Here, we introduce a website that aggregates multiple RNA-Seq studies of the pineal gland spanning years of transcriptomics research on six species across three vertebrate classes, processed uniformly and presented in a user-friendly site allowing inspection of individual genes as well as UCSC Genome Browser track hubs of each experiment. The site can be found at <https://snengs.nichd.nih.gov>. The results of this effort facilitate the comparative and evolutionary analysis of the pineal gland and retina, reflecting an interest in the evolutionary history that links these tissues as derivatives of a common ancestral photodetector<sup>21,22</sup>. Most of the tens of thousands of transcripts profiled are otherwise absent from the pineal and retinal literature and in many cases have not been well studied in any tissue. Accordingly, the web page opens new avenues of research.

This report alerts investigators to the availability of this resource, which will be of special value where user-friendly compiled pineal and retinal RNA-Seq data are otherwise not available in any format. The graphs and other information extracted from the web page are in the public domain. The web page and its underlying infrastructure is designed to be easily updated as data from new experiments become available or as reanalysis of existing datasets using improved software and updated genomes is completed.

## Methods

### Animal

Samples were collected to identify differential day/night ratios; and, in the case of the rat, expression was studied as a function of development, denervation and adrenergic–cyclic AMP stimulation (Table 1) <sup>9,12,23–25</sup>. In many cases retinal tissue was profiled in parallel. Mixed tissue RNA samples were used in conjunction with the pineal gland and retina to estimate the enrichment of a transcript.

### Sequencing

Illumina sequencing was used in all cases. Specific experimental details are available on each experiment's page on the website (see <https://snengs.nichd.nih.gov/experiments>), that recapitulates original experimental methods in the respective original manuscript. All data across all species were re-analyzed in the assemblies and annotations described (<https://snengs.nichd.nih.gov/methods>). Quality of FASTQ files were analyzed with FastQC v0.11.8 and MultiQC v1.6 and all samples demonstrated high-quality sequencing. Adapters were

removed and light quality trimming was performed with cutadapt v1.18 using additional arguments -q 20 --minimum-length 25.

These trimmed reads were provided to Salmon v0.12.0<sup>26</sup> for transcript quantification using an index built for transcriptomes as described in (<https://snengs.nichd.nih.gov/methods>), and run using the additional arguments -gcBias -seqBias --libTypeA. For each gene, the per-transcript values reported by Salmon v0.12.0 were summed to provide a gene-level expression estimate in units of transcripts per million reads (TPM). These are the values reported in the tables and plots of each gene page.

### Genomic visualization

For genomic visualization in the UCSC genome browser, trimmed reads were aligned using HISAT2 v2.1.0 to the respective genome indicated below. From these aligned reads, normalized bigWig files were created using the deepTools v3.1.3 bamCoverage tool using additional options --minMappingQuality 20 --smoothLength 10 --normalizeUsing BPM --binSize 1 such that multi-mappers were ignored. For stranded libraries, the tool was run twice: once with --filterRNAstrand forward and once with --filterRNAstrand reverse to get separate tracks for each strand. The resulting bigWig files were combined into a UCSC track hub using the trackhub Python package.

UCSC (typically used for extensive visualization capabilities) and Ensembl (typically used for its comprehensive annotations) are not consistent in their chromosome nomenclature. To facilitate linking from gene-level transcription estimates on this website to genomic signal at UCSC, we converted chromosome names from Ensembl to the UCSC equivalents by matching the md5sums of each chromosome; see GitHub repository (<https://github.com/NICHD-BSPC/chrom-name-mappings>) for details and code.

### Genome and transcriptome assemblies

For each species, the genomic assembly indicated (<https://snengs.nichd.nih.gov/methods>) was used for visualization in the UCSC Genome Browser, while the transcriptome was used to calculate TPM expression estimates to display in plots on individual gene pages.

### Implementation details

The website is written in the Python programming language using the ‘Flask’ framework. Configuration of the website is driven by a YAML format file that points to Salmon v0.12.0 output along with details like methods descriptions, UCSC track hub colors, bar plot colors, and any other experiment-specific configuration. This greatly streamlines the process of adding new studies and new species. Data were processed using lcdb-wf (<https://github.com/lcdb/lcdb-wf>), which is itself driven by YAML configuration in a species-agnostic manner, allowing for uniform processing across all studies.

## Results

The Home Page (<https://snengs.nichd.nih.gov>) introduces the user to the main sections of the database (Figure 1A). Selecting the Search section displays the Search subpage (Figure 1B).

Entering a gene symbol (*i.e.*, Aanat) in the query box opens the Results subpage, which contains a listing of species and experiments (Figure 2). The Search function will accept alternative symbols; however, when difficulty is encountered obtaining a result, the user is encouraged to refer to gene databases for assistance. This page contains information on the samples, including species, a brief description of the experiment and a Link to the gene page.

Clicking on that Link (Figure 2), opens a Gene subpage (Figure 3) with links to the Ensembl data (gene id) and the UCSC Genome Browser for the gene (Open UCSC track for this gene), in addition to presenting experimental results in a bar graph. These results are normalized count data (in TPM). In cases where multiple experiments for a species exist, all experiments are displayed and can be viewed by scrolling vertically.

Selecting Experiments from the Home page displays the experiments with four links (Figure 4). The first is the Search subpage, described above. The second (Download) retrieves the data in an Excel file. The third retrieves the Details (Figure 5) of the experiment, including sample preparation and data analysis; scrolling horizontally is necessary to open the table. The last is a Link to the UCSC Genome Browser, which documents the location of reads mapping for each gene.

The Methods and Help pages are not presented as figures. The Methods page contains general information on the Bioinformatics methods and identifies the genomic assemblies used; the Help subpage has links to tutorials on the use of the UCSC browser and contact information for further assistance.

As an example of the utility of comparing data across multiple species in a uniform format, we searched for differences in the day/night levels of transcripts among species. As shown in Table 2, the large night/day rhythms in the transcript abundance of several genes in the rat are not seen in the rhesus monkey or to a similar degree in other species (<https://snengs.nichd.nih.gov/search>). This emphasizes the importance of posttranslational modifications that occur<sup>27,28</sup>. It also is a caution against making generalities based on studies of one species.

The data also focus on the similarity of the genomic profiles of pineal glands from the species studied (Table 3). Selective expression of each gene was calculated as the ratio of expression of a specific gene in the pineal gland to that in a mixture of RNA from a group of tissues. As expected, three genes responsible for melatonin synthesis (Tph1, Aanat and Asmt) were selectively expressed in the glands studied. Another group of genes selectively expressed in the pineal gland includes those established as markers of the retina. The high expression of these genes only in the pineal gland and retina is known<sup>21,22</sup>. However, the specific functions of these retina-related genes and other selectively expressed genes in the pineal gland have not received significant attention and deserve further analysis.

An analysis of the conserved highly expressed and tissue specific transcripts in the pineal gland, retina and in both tissues (Table 4) was done by identifying the highly tissue specific transcripts. They were then binned according to their expression ratio (pineal gland: retina). The results reveal a relatively smaller sets of pineal-specific and retina-specific genes, and a

larger group of genes expressed in both tissues. Noting that these genes are selectively expressed only in these two tissues and not in others, it is highly likely that these genes represent evolutionarily conserved elements that can be considered to be related to the common origin of both tissues. In some cases, their roles have been identified, but in many cases, a functional role has not been established.

## Discussion

This database will serve as a foundation for future molecular biological research on the pineal gland and retina, making available the data to scientists with a computer and an internet connection. The uniform processing of raw data makes the comparison of results more meaningful and takes advantage of advances in tools, algorithms, assemblies, and annotations since original publication. Whereas the human and mouse genomes are the most highly annotated, and the chicken and zebrafish less so, the maturity of all annotations allows for in-depth analysis of nearly all genes. A potential problem is that symbols used for identification of a gene in one species may not be used in other species or may be used for different genes. Hence, in cases where identification is questionable, confirmation may require analysis of sequence homology.

### Utility and accuracy of RNA-Seq data

In judging the utility and accuracy of the RNA-Seq data, it should be noted that there is good agreement with data from other methods for the analysis of pineal gland and retina material, including microarray, Northern blot and qRT-PCR as regards day/night differences<sup>5–13</sup>. Accordingly, the RNA-Seq data can be viewed as highly useful and reliable.

The method also has advantages over other methods, perhaps the most important is that it sequences all transcripts, including those without a history in any literature. This opens new avenues for study for the pineal gland and retina. One of the most fertile areas is the identification of noncoding RNAs, both micro RNAs and long noncoding RNAs<sup>9,23</sup>. Some of these are known to have daily rhythms in both tissues. Noteworthy is the discovery of a unique micro RNA-183–96-182 cluster in the pineal gland and retina<sup>9,29</sup>, which represents the major component of pineal miRNAs. Accordingly, it can be considered to be an additional marker of the common ancestral photodetector which gave rise to the pineal gland and retina. Although the function of this cluster remains unknown in the pineal gland, it has been reported to play a role in phototransduction and development in the eye<sup>30,31</sup>.

Study of pineal miRs also led to the discovery of very high levels of pY RNA1-s2 in the retina, relative to other tissues<sup>25</sup>, including the pineal gland. Moreover, it was found that pY RNA1-s2 selectively binds the nuclear matrix protein Matrin 3 and to a lesser degree to heterogeneous nuclear ribonucleoprotein U-like protein. The distribution of py RNA1-s2 in all retinæ and retinal cell lines suggests a role in vision. Both these discoveries could not have been made using methods other than RNAseq.

Likewise, the finding of robust daily rhythms in the abundance of several long non-coding RNAs<sup>23</sup> in the pineal gland under neural control, and the discovery of expression of lncSN134 in both the retina and pineal gland was dependent on the use of RNAseq. The long



noncoding RNAs range in size significantly and like pineal miRs, remain largely unstudied and unknown.

Whereas RNAseq is a powerful technique, the results must be viewed with healthy skepticism, especially with transcripts that are weakly expressed and when evaluating small night/day differences in transcript abundance. Confirmation by an independent method should be considered. In addition, in the case of weakly expressed transcripts, the mapping of reads on the UCSC browser should be evaluated to confirm that the read assignment pattern is consistent with the intron/exon features of the transcript.

### Experimental design

A problem that is considered in any study designed to measure day/night differences is the number of time points per day. Often this is limited by factors including the housing of animals and the number of animals per point necessary to obtain sound data. RNA-Seq introduces another factor, the cost of analysis. Accordingly, the design of the studies included in the database (Table 1) is also a reflection of the cost of sequencing and bioinformatics. The studies included sampling that ranged from two to six time points per day. When sampling is done at only two time points, noon and midnight, the potential for overlooking a dawn/dusk rhythm exists. Accordingly, it is best not to limit experiments to two time point studies, and to use four or more to detect daily rhythms. However, in the case of study of daily rhythms in the pineal gland, a two time point study will capture most large changes. Moreover, this approach is highly instructive, in that it provides valuable data on the levels of tens of thousands of transcripts. Accordingly, one can see merit in such studies.

The number of replicates to use is also another important issue. RNA-seq data are typically highly reproducible for most transcripts when normalized. This reflects a feature of the method, in that there is redundancy in the detection of a transcript, as a result of fragmentation and amplification. In the final analysis, each calculated transcript level is not simply a single measurement, but reflects multiple detection events, depending on the size of the transcript and abundance. Accordingly, in N=1 situations, it is possible to obtain an indication of statistical variance of all transcripts, and use this to determine whether, for example, a day/night difference is statistically significant.

### Transcriptomics versus proteomics

Whereas RNA-Seq does provide a highly useful tool for the discovery and characterization of transcripts, it is not a substitute for proteomics. The study of an mRNA and its encoded protein often are in agreement as regards the presence and dynamic changes in both. However, this is clearly not the case in all situations.

An excellent example is Aanat. In the rat, Aanat mRNA, protein and activity increase at night, reflecting phosphorylation of the protein at two sites. When lights are turned on in the middle of the night, a rapid decrease in enzyme activity occurs, with little change in mRNA levels. The changes in enzyme activity are due to dephosphorylation of the protein, which is rapidly destroyed by proteasomal proteolysis, as reviewed<sup>32</sup>. Another example of mRNA levels and protein levels not exhibiting similar dynamics is found in studies of the rhesus pineal gland. There is little daily change in mRNA encoding Aanat, although the changes in

enzyme activity are robust<sup>33</sup>. These observations are evidence that it is necessary to determine whether changes in an mRNA are associated with changes in an encoded protein to determine the relationship. Unfortunately, the science of proteomics has not advanced to the all-inclusive nature of mRNA analysis, in part because it is difficult to uniformly detect the possible posttranslational modifications.

Use of the database will allow investigators to initiate efforts to identify transcripts that are highly expressed in the pineal gland relative to the retina and or other tissues, transcripts that are highly expressed in the pineal gland of one species but not another, transcripts that exhibit marked night/day differences, transcripts that are under neural/adrenergic cyclic AMP control, and transcripts that exhibit changes in expression during development. In doing so, the web page should promote and enhance future studies of pineal cell biology.

### Referencing the web page

The data on the web page are in the public domain and the use of the figures and data does not require authorization of the authors. The web page should be referenced by citing this publication.

### Acknowledgements

The authors wish to express their appreciation for discussion and testing services provided by Apratim Mitra and Sydney Hertafeld of the Bioinformatics and Scientific Programming Core, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD). The authors also want to acknowledge the contributions of the following: David "Dr. J." Jacobowitz, Collaborative Health Initiative Research Program, Uniformed Services University of the Health Sciences; Stephen W. Hartley and James C. Mullikin, National Human Genome Research Institute, National Institutes of Health; Leming Shi, United States Food and Drug Administrations, National Center for Toxicological Research and the School of Basic Medical Sciences, Anhui Medical University, Hefei; and, Artem Zykovich, NICHD. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

### References

- (1). Darmon MC; Guibert B; Leviel V; Ehret M; Maitre M; Mallet J Sequence of two mRNAs encoding active rat tryptophan hydroxylase. *Journal of neurochemistry* 1988, 51, 312–316. [PubMed: 3379411]
- (2). Grenett HE; Ledley FD; Reed LL; Woo SL Full-length cDNA for rabbit tryptophan hydroxylase: functional domains and evolution of aromatic amino acid hydroxylases. *Proc Natl Acad Sci U S A* 1987, 84, 5530–5534. [PubMed: 3475690]
- (3). Ishida I; Obinata M; Deguchi T Molecular cloning and nucleotide sequence of cDNA encoding hydroxyindole O-methyltransferase of bovine pineal glands. *J Biol Chem* 1987, 262, 2895–2899. [PubMed: 3818627]
- (4). Dumas S; Darmon MC; Delort J; Mallet J Differential control of tryptophan hydroxylase expression in raphe and in pineal gland: evidence for a role of translation efficiency. *J Neurosci Res* 1989, 24, 537–547. [PubMed: 2600977]
- (5). Zilberman-Peled B; Bransburg-Zabary S; Klein DC; Gothilf Y Molecular evolution of multiple arylalkylamine N-acetyltransferase (AANAT) in fish. *Mar Drugs* 2011, 9, 906–921. [PubMed: 21673898]
- (6). Appelbaum L; Toyama R; Dawid IB; Klein DC; Baler R; Gothilf Y Zebrafish serotonin-N-acetyltransferase-2 gene regulation: pineal-restrictive downstream module contains a functional E-box and three photoreceptor conserved elements. *Mol Endocrinol* 2004, 18, 1210–1221. [PubMed: 14988431]



- (7). Ben-Moshe Livne Z; Alon S; Vallone D; Bayleyen Y; Tovin A; Shainer I; Nisembaum LG; Aviram I; Smadja-Storz S; Fuentes M; Falcon J; Eisenberg E; Klein DC; Burgess HA; Foulkes NS; Gothilf Y Genetically Blocking the Zebrafish Pineal Clock Affects Circadian Behavior. *PLoS Genet* 2016, 12, e1006445. [PubMed: 27870848]
- (8). Hartley SW; Coon SL; Savastano LE; Mullikin JC; Program NCS; Fu C; Klein DC Neurotranscriptomics: The Effects of Neonatal Stimulus Deprivation on the Rat Pineal Transcriptome. *PLoS One* 2015, 10, e0137548. [PubMed: 26367423]
- (9). Clokie SJ; Lau P; Kim HH; Coon SL; Klein DC MicroRNAs in the pineal gland: miR-483 regulates melatonin synthesis by targeting arylalkylamine N-acetyltransferase. *J Biol Chem* 2012, 287, 25312–25324. [PubMed: 22908386]
- (10). Rovsing L; Clokie S; Bustos DM; Rohde K; Coon SL; Litman T; Rath MF; Moller M; Klein DC Crx broadly modulates the pineal transcriptome. *Journal of neurochemistry* 2011, 119, 262–274. [PubMed: 21797868]
- (11). Bustos DM; Bailey MJ; Sugden D; Carter DA; Rath MF; Moller M; Coon SL; Weller JL; Klein DC Global daily dynamics of the pineal transcriptome. *Cell Tissue Res* 2011, 344, 1–11. [PubMed: 21302120]
- (12). Backlund PS; Urbanski HF; Doll MA; Hein DW; Bozinoski M; Mason CE; Coon SL; Klein DC Daily Rhythm in Plasma N-acetyltryptamine. *J Biol Rhythms* 2017, 32, 195–211. [PubMed: 28466676]
- (13). Tovin A; Alon S; Ben-Moshe Z; Mracek P; Vatine G; Foulkes NS; Jacob-Hirsch J; Rechavi G; Toyama R; Coon SL; Klein DC; Eisenberg E; Gothilf Y Systematic identification of rhythmic genes reveals camk1gb as a new element in the circadian clockwork. *PLoS Genet* 2012, 8, e1003116. [PubMed: 23284293]
- (14). Bailey MJ; Beremand PD; Hammer R; Bell-Pedersen D; Thomas TL; Cassone VM Transcriptional profiling of the chick pineal gland, a photoreceptive circadian oscillator and pacemaker. *Mol Endocrinol* 2003, 17, 2084–2095. [PubMed: 12881511]
- (15). Karaganis SP; Kumar V; Beremand PD; Bailey MJ; Thomas TL; Cassone VM Circadian genomics of the chick pineal gland in vitro. *BMC Genomics* 2008, 9, 206. [PubMed: 18454867]
- (16). Collado-Torres L; Nellore A; Kammers K; Ellis SE; Taub MA; Hansen KD; Jaffe AE; Langmead B; Leek JT Reproducible RNA-seq analysis using recount2. *Nat Biotechnol* 2017, 35, 319–321. [PubMed: 28398307]
- (17). Kent WJ; Sugnet CW; Furey TS; Roskin KM; Pringle TH; Zahler AM; Haussler D The human genome browser at UCSC. *Genome Res* 2002, 12, 996–1006. [PubMed: 12045153]
- (18). Raney BJ; Dreszer TR; Barber GP; Clawson H; Fujita PA; Wang T; Nguyen N; Paten B; Zweig AS; Karolchik D; Kent WJ Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 2014, 30, 1003–1005. [PubMed: 24227676]
- (19). Roberts GP; Larraufie P; Richards P; Kay RG; Galvin SG; Miedzybrodzka EL; Leiter A; Li HJ; Glass LL; Ma MKL; Lam B; Yeo GSH; Scharfmann R; Chiarugi D; Hardwick RH; Reimann F; Gribble FM Comparison of Human and Murine Enteroendocrine Cells by Transcriptomic and Peptidomic Profiling. *Diabetes* 2019, 68, 1062–1072. [PubMed: 30733330]
- (20). Kansler ER; Verma A; Langdon EM; Simon-Vermot T; Yin A; Lee W; Attiyeh M; Elemento O; White RM Melanoma genome evolution across species. *BMC Genomics* 2017, 18, 136. [PubMed: 28173755]
- (21). Klein DC Evolution of the vertebrate pineal gland: the AANAT hypothesis. *Chronobiol Int* 2006, 23, 5–20. [PubMed: 16687276]
- (22). Klein DC The 2004 Aschoff/Pittendrigh lecture: Theory of the origin of the pineal gland--a tale of conflict and resolution. *J Biol Rhythms* 2004, 19, 264–279. [PubMed: 15245646]
- (23). Coon SL; Munson PJ; Cherukuri PF; Sugden D; Rath MF; Moller M; Clokie SJ; Fu C; Olanich ME; Rangel Z; Werner T; Mullikin JC; Klein DC Circadian changes in long noncoding RNAs in the pineal gland. *Proc Natl Acad Sci U S A* 2012, 109, 13319–13324. [PubMed: 22864914]
- (24). Yamazaki F; Moller M; Fu C; Clokie SJ; Zykovich A; Coon SL; Klein DC; Rath MF The Lhx9 homeobox gene controls pineal gland development and prevents postnatal hydrocephalus. *Brain structure & function* 2014, 220, 1497–1509. [PubMed: 24647753]

- (25). Yamazaki F; Kim HH; Lau P; Hwang CK; Iuvone PM; Klein D; Clokie SJ pY RNA1-s2: a highly retina-enriched small RNA that selectively binds to Matrin 3 (Matr3). *PLoS One* 2014, 9, e88217. [PubMed: 24558381]
- (26). Patro R; Duggal G; Love MI; Irizarry RA; Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017, 14, 417–419. [PubMed: 28263959]
- (27). Ganguly S; Gastel JA; Weller JL; Schwartz C; Jaffe H; Namboodiri MA; Coon SL; Hickman AB; Rollag M; Obsil T; Beauverger P; Ferry G; Boutin JA; Klein DC Role of a pineal cAMP-operated arylalkylamine N-acetyltransferase/14–3–3-binding switch in melatonin synthesis. *Proc Natl Acad Sci U S A* 2001, 98, 8083–8088. [PubMed: 11427721]
- (28). Ganguly S; Weller JL; Ho A; Chemineau P; Malpaux B; Klein DC Melatonin synthesis: 14–3–3-dependent activation and inhibition of arylalkylamine N-acetyltransferase mediated by phosphoserine-205. *Proc Natl Acad Sci U S A* 2005, 102, 1222–1227. [PubMed: 15644438]
- (29). Xu S; Witmer PD; Lumayag S; Kovacs B; Valle D MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. *J Biol Chem* 2007, 282, 25053–25066. [PubMed: 17597072]
- (30). Li H; Gong Y; Qian H; Chen T; Liu Z; Jiang Z; Wei S Brain-derived neurotrophic factor is a novel target gene of the has-miR-183/96/182 cluster in retinal pigment epithelial cells following visible light exposure. *Mol Med Rep* 2015, 12, 2793–2799. [PubMed: 25955435]
- (31). Xiang L; Chen XJ; Wu KC; Zhang CJ; Zhou GH; Lv JN; Sun LF; Cheng FF; Cai XB; Jin ZB miR-183/96 plays a pivotal regulatory role in mouse photoreceptor maturation and maintenance. *Proc Natl Acad Sci U S A* 2017, 114, 6376–6381. [PubMed: 28559309]
- (32). Klein DC Arylalkylamine N-acetyltransferase: “the Timezyme”. *J Biol Chem* 2007, 282, 4233–4237. [PubMed: 17164235]
- (33). Coon SL; Del Olmo E; Young WS 3rd; Klein DC Melatonin synthesis enzymes in *Macaca mulatta*: focus on arylalkylamine N-acetyltransferase (EC 2.3.1.87). *J Clin Endocrinol Metab* 2002, 87, 4699–4706. [PubMed: 12364461]

A.

SNE-NICHD Transcriptome Profiling

Home Search Experiments Methods Help

## Home

The SNE-NICHD Transcriptome Profiling web page makes RNA-seq data from the NICHD Section on Neuroendocrinology accessible to investigators interested in dynamic changes in the pineal transcriptome.

Data has been drawn from studies on the zebrafish, chicken, mouse, rat, rhesus macaque, and human. In most cases samples have been obtained at midday and midnight; in some cases more frequent sampling has been employed. In addition, data from developmental studies on the rat are available.

Raw data from all experiments have been completely reanalyzed using the latest available assemblies and annotations and using a uniform processing pipeline for consistent results (see [bioinformatics methods](#)).

Processed data are available in several forms:

- download as Excel spreadsheets for all genes, from the [experiment details page](#)
- view as bar plots on individual gene pages (see, for example the page for [human AANAT](#), or [search](#) for a gene of interest)
- browse genomic signal in track hubs on the UCSC Genome Browser. For example, the human AANAT signal across all samples can be viewed [here](#).

Figures generated by the web page are in the public domain and can be used freely, citing the URL of the web page.

B.

## Search

Enter (case-insensitive) gene symbol or Ensembl ID. Example: Aanat



### Figure 1. Home page and Search subpage.

1A. The Home page (<https://snengs.nichd.nih.gov/home>) opens subpages which are organized to search for genes specifically and to retrieve information relevant to the entire dataset. In addition, the Methods page contains useful information about the experiments and the Help page has useful videos on use of the UCSC Genome Browser. 1B. The Search page (<https://snengs.nichd.nih.gov/search>). Entering an official gene symbol or an Ensemble ID symbol in the Search box retrieves data from all species. For aliases please refer to the Ensembl or NCBI Gene databases.

## Results

Columns are sortable. Use the text field to the right to search within this subset of results (e.g., to restrict results to a particular species). The "View" link opens a new page with details on that gene across multiple conditions of the experiment.

Search within these results:

| Searched text               | Species   | Experiment name  | Link to gene page    |
|-----------------------------|-----------|--|----------------------|
| AANAT   ENSG00000129673     | Human     | Pineal gland; day and night  | <a href="#">View</a> |
| AANAT   ENSGALG00000001955  | Chicken   | Pineal gland and retina; time series; constant darkness  | <a href="#">View</a> |
| AANAT   ENSMMUG00000018970  | Rhesus    | Pineal gland, retina and mixed tissue; time series   | <a href="#">View</a> |
| Aanat   ENSMUSG00000020804  | Mouse     | Pineal gland and retina, retina and mixed tissue; day and night; Eya2 KO                       | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland; day and night; and mixed tissue, day; polyA                                      | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland development; day and night  | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland (RP), retina (RR) and mixed tissue (RX); 24-hr time series                        | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland; day and night; and mixed tissue, day; Ribominus                                  | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland; superior cervical decentralization (DCN) or ganglionectomy (SCGX); day and night | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland in vitro; norepinephrine (NE) or dibutyl cyclic AMP (DBcAMP)                      | <a href="#">View</a> |
| Aanat   ENSRNOG00000011182  | Rat       | Pineal gland marker genes; day and night   | <a href="#">View</a> |
| aanat1   ENSDARG00000036567 | Zebrafish | Pineal gland; time series; constant darkness   | <a href="#">View</a> |
| aanat1   ENSDARG00000036567 | Zebrafish | Eye, pineal gland and mixed tissue; day and night  | <a href="#">View</a> |
| aanat2   ENSDARG00000079802 | Zebrafish | Pineal gland; time series; constant darkness   | <a href="#">View</a> |
| aanat2   ENSDARG00000079802 | Zebrafish | Eye, pineal gland and mixed tissue; day and night  | <a href="#">View</a> |

Showing 1 to 15 of 15 entries

### Figure 2. Results subpage.

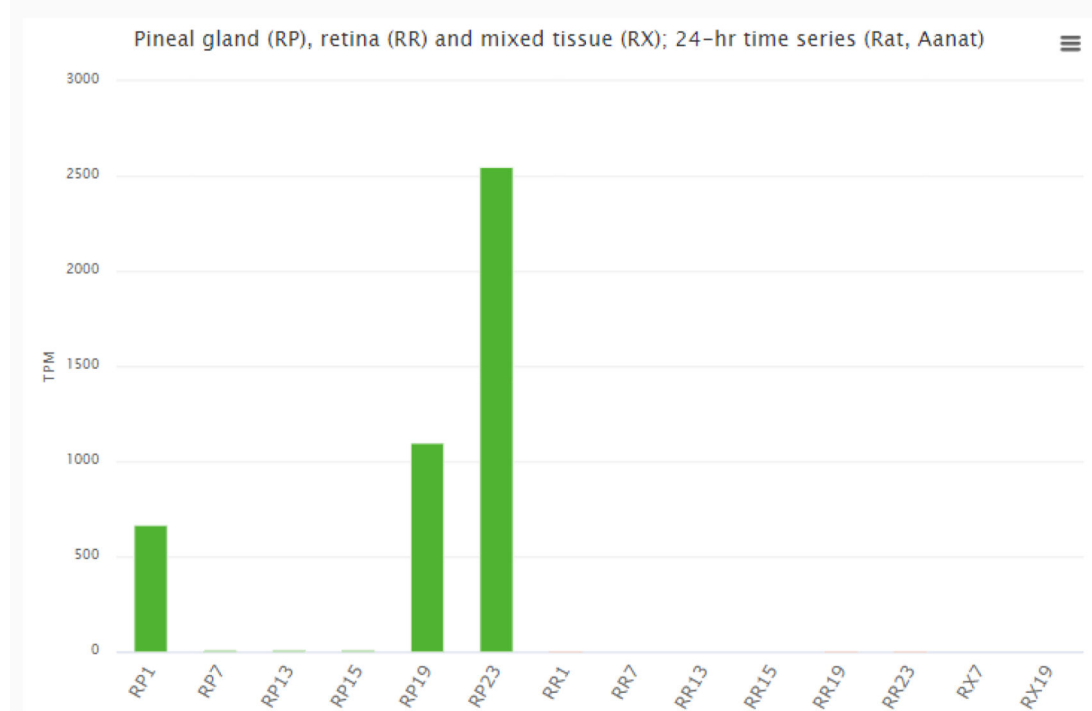
The results of querying a gene symbol generates a listing of the experiments and species in which the gene was found (<https://snengs.nichd.nih.gov/search>). Depending on the size of the gene family, multiple gene symbols may be displayed. In this case, one has to use the data cautiously. From this page, highlighted links (View) direct the user to the Gene subpage, which lists results of a single species and experiment.

# Pineal gland (RP), retina (RR) and mixed tissue (RX); 24-hr time series

## Rat

[\(experiment details\)](#)

|  |                                    |
|--|------------------------------------|
| <b>gene</b>  | ENSRNOG00000011182                 |
| <b>gene name</b>   | Aanat                              |
| <b>gene biotype</b>  | protein_coding                     |
| <b>chrom</b>   | chr10                              |
| <b>mean</b>  | 313.2                              |
| <b>stddev</b>  | 722.8                              |
| <b>gene id</b>   | <a href="#">ENSRNOG00000011182</a> |
| <b>coordinates</b>   | chr10:105568091-105572403          |
| <a href="#">Open UCSC track hub for this gene</a> (see help) |                                    |



### Figure 3. Gene subpage.

An example of the gene page ([https://snengs.nichd.nih.gov/species/Rat/gene/ENSRNOG00000011182#Time\\_Series-anchor](https://snengs.nichd.nih.gov/species/Rat/gene/ENSRNOG00000011182#Time_Series-anchor)) that displays information about a specific gene including relevant experimental details, the UCSC track hub (Open UCSC track hub for this gene) and a help page for use of the UCSC Genome Browser. General information about each gene is available by clicking on the gene id, e.g. ENSRNOG00000011182. Experimental results are displayed below in a bar graph. In addition, accessing the results of a single experiment will open other experiments dependent on availability.

## Experiments

This page summarizes the experiments across multiple species and multiple experiments.

- The **Data** column contains a link to an Excel file for that experiment. Values are in TPM (Transcripts Per Million) estimates, see [Bioinformatics methods](#) for details.
- The **Details** column links to a page detailing the experimental methods. It includes a list of GEO and SRA accessions used.
- The **UCSC track hub** column links to the UCSC genome browser which displays all tracks for the experiment into a configurable track hub.

| Species   | Experiment name  | Search                 | Data                     | Details                 | UCSC Track Hub       |
|-----------|--|------------------------|--------------------------|-------------------------|----------------------|
| Chicken   | Pineal gland and retina; time series; constant darkness  | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Human     | Pineal gland; day and night  | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Mouse     | Pineal gland and retina, retina and mixed tissue; day and night; Eya2 KO                       | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland; day and night; and mixed tissue, day; polyA                                      | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland development; day and night  | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland (RP), retina (RR) and mixed tissue (RX); 24-hr time series                        | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland; day and night; and mixed tissue, day; Ribominus                                  | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland; superior cervical decentralization (DCN) or ganglionectomy (SCGX); day and night | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland in vitro; norepinephrine (NE) or dibutyl cyclic AMP (DbcAMP)                      | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rat       | Pineal gland marker genes; day and night   | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Rhesus    | Pineal gland, retina and mixed tissue; time series   | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Zebrafish | Pineal gland; time series; constant darkness   | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |
| Zebrafish | Eye, pineal gland and mixed tissue; day and night  | <a href="#">Search</a> | <a href="#">Download</a> | <a href="#">Details</a> | <a href="#">Link</a> |

### Figure 4. Experiments subpage.

The Experiments subpage (<https://snengs.nichd.nih.gov/experiments>) is accessed from the Home page. It is an index for all experiments, leading to several resources. The highlighted link retrieves the Search subpage, described above. The Data link (Download) returns the data for an entire experiment in an Excel file, which also contains the average expression values and variance. Selecting Details opens the page with experimental details (see Figure 5). The highlighted link to the UCSC Track Hub (Link) gives access to the mapped data on the UCSC Genome Browser.



## Pineal gland (RP), retina (RR) and mixed tissue (RX); 24-hr time series (Rat)

### Specimens

Pools (>10 animals) of mixed sexes rat pineal glands or retinas were prepared from tissue obtained at six time points during a 24-hour period. Animals were housed in a 14:10 light:dark lighting cycle. Samples were collected at mid-day (ZT7) and mid-night (ZT19), and at one hour before and one hour after the light-to-dark transition (ZT13 and ZT15) and the dark-to-light transition (ZT23 and ZT1). In addition, a mixed tissue sample was prepared from material collected from three rats at ZT7 and at ZT19. The tissues were: cortex, cerebellum, midbrain, hypothalamus, hindbrain, spinal cord, retina, pituitary, heart, liver, lung, kidney, skeletal muscle, small intestine, adrenal gland.

### RNA Preparation

Samples of tissue from the three animals were mixed. Total RNA was extracted using an RNeasy Kit with on-column DNase treatment. To prepare the mixed tissue sample, equal amounts of RNA from each of the 15 tissues were combined.

### RNA-Seq library preparation and sequencing

Total RNA (10 ug) was polyA-selected, fragmented and barcoded. Samples were run six per sequencing lane on an Illumina HiSeq2000 machine, generating paired-end 101-mer reads.

**Note:** In some cases, the mixed tissue values for selected genes are exceptionally high because that gene is very highly expressed in only one tissue, e.g., Rho and GH are strongly expressed only in the retina and pituitary gland, respectively.

### Bioinformatics methods

See [bioinformatics methods](#) for details.

## Samples

The following table describes the sample metadata for this experiment. Accessions in GEO, BioProject, and SRA link to their respective entries. Raw data were downloaded as FASTQ format using *fastq-dump* for each of the SRR run accessions.

Search:

| Sample <sup>▲</sup> | Label <sup>⚙</sup> | BioSample <sup>⚙</sup>       | Experiment <sup>⚙</sup>   | SRRrun <sup>⚙</sup>       | SRA_Sample <sup>⚙</sup>   | Sample_Name <sup>⚙</sup> | SRA_Study                 |
|---------------------|--------------------|------------------------------|---------------------------|---------------------------|---------------------------|--------------------------|---------------------------|
| RP13_ad             | RP13               | <a href="#">SAMN02045889</a> | <a href="#">SRX265397</a> | <a href="#">SRR827500</a> | <a href="#">SRS413246</a> | GSM1122815               | <a href="#">SRP021090</a> |
| RP15_ad             | RP15               | <a href="#">SAMN02045890</a> | <a href="#">SRX265398</a> | <a href="#">SRR827501</a> | <a href="#">SRS413248</a> | GSM1122816               | <a href="#">SRP021090</a> |
| RP19_ad             | RP19               | <a href="#">SAMN02045891</a> | <a href="#">SRX265399</a> | <a href="#">SRR827502</a> | <a href="#">SRS413247</a> | GSM1122817               | <a href="#">SRP021090</a> |

### Figure 5. Details subpages.

The Details subpages are accessed from the Experiments subpage (see Figure 4; <https://snengs.nichd.nih.gov/experiments>) by clicking on Details for a specific experiment. This yields information on sample preparation, RNA preparation, and data processing; and, the location of archived data. The search box is used to interrogate the table with identifiers (e.g. SRX3229487) or fragments of identifiers (e.g. \_04h) in the table. The Samples section in this figure is truncated for presentation purposes.

**Table 1.**

Experiments on the database.

| Study No. | Animals                | Experiment name  | Tissue      | Lighting  | Sampling times          | Notes                                | Reps | Refs   |
|-----------|------------------------|--|-------------|-----------|-------------------------|--------------------------------------|------|--------|
| 101       | Chicken, White Leghorn | Pineal gland and retina; time series; constant darkness  | PG, R       | D:D       | CT 0, 4, 8, 12, 16, 20  | N/A                                  | 3    | N/A    |
| 102       | Human                  | Pineal gland; day and night  | PG          | L:D 12:12 | ZT 6, 18                | N/A                                  | 2, 4 | N/A    |
| 103       | Mouse, 129sv           | Pineal gland, retina and mixed tissue; day and night; Eya2 KO                                  | PG, R, MT   | L:D 12:12 | ZT 6, 18                | Eya2 KO                              | 1    | N/A    |
| 104       | Rat, Sprague Dawley    | Pineal gland; day and night; and mixed tissue, day; polyA                                      | PG, MT      | L:D 14:10 | ZT 7, 19                | N/A                                  | 1    | 23     |
| 105       | Rat, Sprague Dawley    | Pineal gland development; day and night  | PG          | L:D 14:10 | ZT 7, 19                | Ages: E21, P5, P20, P40              | 1    | 23, 24 |
| 106       | Rat, Sprague Dawley    | Pineal gland (RP), retina (RR) and mixed tissue (RX); 24-hr time series                        | PG, R, MT   | L:D 14:10 | ZT 1, 7, 13, 15, 19, 23 | N/A                                  | 1    | 23     |
| 107       | Rat, Sprague Dawley    | Pineal gland; day and night; and mixed tissue, day; Ribominus                                  | PG, MT      | L:D 14:10 | ZT 7, 19                | N/A                                  | 1    | 23     |
| 108       | Rat, Sprague Dawley    | Pineal gland; superior cervical decentralization (DCN) or ganglionectomy (SCGX); day and night | PG          | L:D 14:10 | ZT 7, 19                | DCN, SCGX, Sham, Control             | 3    | 8      |
| 109       | Rat, Sprague Dawley    | Pineal gland in vitro; norepinephrine (NE) or dibutyl cyclic AMP (DBcAMP)                      | PG          | N/A       | N/A                     | Cultured glands; NE, DBcAMP, Control | 3    | 8      |
| 110       | Rat, Sprague Dawley    | Pineal gland marker genes; day and night   | PG          | L:D 14:10 | ZT 7, 19                | N/A                                  | 1    | 23     |
| 111       | Rhesus macaque         | Pineal gland, retina and mixed tissue; time series   | PG, R, MT   | L:D 12:12 | ZT 6, 12, 18, 24        | Dawn, Day, Dusk, Night               | 3    | 12     |
| 112       | Zebrafish              | Pineal gland; time series; constant darkness   | PG          | D:D       | CT 2, 6, 10, 14, 18, 22 | N/A                                  | 2    | 7      |
| 113       | Zebrafish              | Eye, pineal gland and mixed tissue; day and night  | Eye, PG, MT | L:D 12:12 | ZT 6, 18                | clocka KO                            | 1    | N/A    |

Thirteen experiments encompassing six species are on the website; additional experiments are to be added as data become available. Experimental details are available on the website (<https://snengs.nichd.nih.gov/experiments>) and the listed references. Abbreviations: CT, circadian time; DD, constant darkness; L:D, light:dark; MT, mixed tissue; N/A, not available; PG, pineal gland; R, retina; Reps, number of replicates; Refs, references; RP, rat pineal gland; RR, rat retina; RX rat mixed tissue; ZT, Zeitgeber time.

**Table 2.**

Comparative analysis of rhythmic transcript levels in the vertebrate pineal gland.

| Species          | Night/Day                    |  | Day/Night                     |  |
|------------------|------------------------------|--|-------------------------------|--|
|                  | >30-fold                     | 3- to 30-fold  | >30-fold                      | 3- to 30-fold  |
| <b>Chicken</b>   | Gos                          | Spes1, Gnb3, Lbh, Lypla1, Prdm8, Aanat, Tph1, Am89a, Ckmt1a, Chga, Ddc, Ndrg   |                               | Rbp4, Rcan2, SSx2ip, Chgb, Calb2, R3hdml, Atoh8, Efr3a   |
| <b>Human</b>     |                              | DUSP1, HKDC1   |                               |  |
| <b>Mouse</b>     | Gh, Prl                      | Aanat, Odc1, Mat2a, Kif5c, Nap115, Tbc1d15, Crem, Tbc1d1, Tjap1, Ndufa3, Syt4, Mitf, Rmdn3, Extl3, Amd1, Ywhaz, Ccn1, Slc3a2, Impa1, Azin1, Prosc, Iqcb1, Crx, Rab3gap1, Srxn1, Manf, Ppa2, Gja1, Psme2, Arf, Cbx7, Tph1, mt-Ts2, Fgf12, Mpp6, Gnai2, Necap1, Tpm4, Atp2a2, Hdhd3, Rnf13, Ip6k1, Dnajb6, Sik3, Ergic1, Tmem229b, Clptm1, Hsph1*, Auh, mt-Tw, mt-Ti                                       | Igkj4, Igkj1, Tpt1-ps3, Ighj4 | Enpp2, Ttr, Chmp1a, Unc119, Ccnd2, Acp2, Atp6v0a2, Tef, Igf2, Ermard, Lamb2, Fabp7, Twf1, Ewsr1, Etf1, Fxyd1, Arih2, Zfand6, Wbscr22, Ndrg1, Tbc1d17, Cox17, Fam166a, Atox1, Rpgrip1, Acker1, mt-T11, Dpysl3, Cisd3, Prpf19, Sag, Tpm3, Ifit46, Apod, Taz                |
| <b>Rat</b>       | Aanat, Atp7b, Slc15a1, Delk3 | Irs2, Crem, Sik1, Ptch1, Cd24, Zrsr1, Rcan1, Kctd3, Bsx, Mat2a, Etnk1, Camk1g, Mbnl2, Gxylt1, Gem, Nptx1, Pcdh1, Eml5, Galnt16, Pde4b, Reep2, Syt4, Tjp2, Snap25, Hbb, Hba-a2, Dnm2, Fkbp5, Man2a1, Fry, Delk1, Mcam, Arhgap24, Hba-a2, Hspa5, Slc17a6, Farp2, Rhob, Cry2, Lamb1, Hsph1, Ncald, Abca1, Mapk6, Ankrd52, Snrk, Slc7a6, Shroom3, Sik2, Ttc8, Nacad, Qsox1, Xpot, Zhx1, Wipf3, Abef1, Frmpd1 | Matr3                         | Gucy1a1, Frmd4b, Eef1a2, Scrn1, Hook1, Ttr, Pdc, Cfl2  |
| <b>Rhesus</b>    |                              | PENK, CCN2, RP1, FAM167a, TGFB3, ATP2A3,   |                               | OPN1SW, VASH1, PDC, GNGT1, LMOD1   |
| <b>Zebrafish</b> | Nr1d1                        | Sik1, Dbpb, Dusp1, Dtx4, Rdh8b, Gjd2b Gpr137bb, Aanat2, CR391986.1, Delk2a, Gucal1a, Tph1a, Gchi1, Ptn, Myh9a, Lpde6ga, Id2b, Cxcl14, Gabarapb   | Nfil3-5                       | Bhlhe40, Rbp3, Cry1aa, Rorcb, Pde6ha, Rorca, Camk1gb, Rbp4, Rp4l, Kera, Ry1bb, Per2, Irbpl, Cyp27c1, Nfil3-6, Pkfb4b, Sagb, Ahcy, Sdha, Eno1a, Add45ga, Tmtops2a, lrp1a, Hbba1, Aldocb, Tmem237b, Gpr146, Aldoa, Jag2b, Aclya, Cry1ba, Ybx1, Rcvrn3, Acadm, Stra6, Hbba1 |

The day and night levels of transcripts were compared by calculating night/day and day/night ratios of normalized values (+ 0.1 TPM). Non-coding RNAs were eliminated. Only the top 1000 genes with official symbols were further grouped by ratios into greater than 30-fold and 3- to 30-fold differences. Genes are listed according to strength of rhythm. Human pineal data are included, noting that times of death and of tissue removal were not tightly controlled; accordingly, the indication of rhythmicity might be impacted. The data were downloaded from the Experiments page. In addition to the single datasets for chicken, human, mouse and rhesus, the zebrafish “eye, pineal gland & mixed tissue” and rat “pineal marker genes” datasets were used.

**Table 3.**

Highly conserved selectively expressed pineal transcripts in four vertebrate species.

|               | Genes selectively expressed among top 1000 genes   |
|---------------|--|
| Three species | Adra1a, Adrb1, Ankrd33, Casz1, Drd4, Gngt <sup>*</sup> , Grk <sup>*</sup> , Grm <sup>*</sup> , Guca1a, Impg1 <sup>*</sup> , Kif <sup>*</sup> , Opn <sup>*</sup> , Pax3, Pcdh <sup>*</sup> , Pla2g <sup>*</sup> , Ppof2, Prph2, Rdh <sup>*</sup> , Rp1 <sup>*</sup> , Rps <sup>*</sup> , Rxrg, Slc16 <sup>*</sup> , Slc6a <sup>*</sup> , Trim <sup>*</sup> .  |
| Four species  | Aanat <sup>*</sup> , Aipl1, Asmt, Bsx, Cabp <sup>*</sup> , Cacna1 <sup>*</sup> , Cacna2d <sup>*</sup> , Celf3, Chrna3 <sup>*</sup> , Chrnb4, Cngb3 <sup>*</sup> , Col <sup>*</sup> , Cplx <sup>*</sup> , Crb <sup>*</sup> , Crx, Gch1, Impg2, Isl2, Kcn <sup>*</sup> , Lhx4, Lrit <sup>*</sup> , Myo <sup>*</sup> , Neurod <sup>*</sup> , Otx2, Pde6 <sup>*</sup> , Ptprn, Rbp3, Slc24 <sup>*</sup> , Slc38a <sup>*</sup> , Tmem <sup>*</sup> , Tph1, Ush2a. |

Genes were ranked according to selective expression in the pineal glands from zebrafish, mouse, rat and rhesus. Expression was normalized (+ 0.1 TPM) and selective expression was calculated relative to expression in a mixture of tissue. The top 1000 selectively expressed genes were identified and those present in three or four out of four of the species are listed below. The data sources are given in the legend to Table 2.

\* Asterisk, more than one homolog exists in some species; for example, Aanat<sup>\*</sup> represents Aanat in mouse, rat and rhesus in addition to Aanat1 and Aanat2 in zebrafish.

**Table 4.**

Transcripts enriched in pineal gland, retina and both the pineal gland and retina.

| Group                          | Enriched transcripts  |  |
|--------------------------------|---|--|
|                                | Four of four species  | Three of four species  |
| <b>Pineal gland</b>            | Aanat <sup>*</sup> , Asmt, Chrb4, Gch1, Gnat2, Gnb3 <sup>*</sup> , Guca1a, Lhx4, Pde6c, Sall1 <sup>*</sup> , Tph1 <sup>*</sup> , Pax3,  | Alx4, Bsx, Chrb3, Gngt2 <sup>*</sup> , Lrrc38, Ptpn20  |
| <b>Pineal gland and retina</b> | Arr3 <sup>*</sup> , Cabp4, Cacna1f <sup>*</sup> , Cacna2d4, Cnga1 <sup>*</sup> , Cngb3, Cplx4 <sup>*</sup> , Crb2 <sup>*</sup> , Crx, Drd4, Fam161a, Gabrr1, Gabrr3 <sup>*</sup> , Gngt1, Grk1 <sup>*</sup> , Guca1b, Impg1 <sup>*</sup> , Impg2 <sup>*</sup> , Kcnb <sup>*</sup> , Lrit1 <sup>*</sup> , Mpp4, Msi1, Myo <sup>*</sup> , Nyx, Opn1sw, Otx2 <sup>*</sup> , Pdc <sup>*</sup> , Pde6g <sup>*</sup> , Rbp3, Rlbp1, Rom1 <sup>*</sup> , Rp111, Slc24a1, Stx3, Tulp <sup>*</sup> , Unc119 <sup>*</sup> , Ush2a | Adrb1, Cabp5 <sup>*</sup> , Crabp <sup>*</sup> , Crb1, Crocc, Egflam, Fabp <sup>*</sup> , Fam169a, Gnb5, Gng1, Grik1 <sup>*</sup> , Gucy2d, Hcn <sup>*</sup> , Igsf9, Impdh1, Kcn <sup>*</sup> , Kcna <sup>*</sup> , Kcnj14, Lrit2, Lrit3, Mak, Mgarp, Neurod4, Ntng2, Nxn11, Pcdh15, Pla2g <sup>*</sup> , Plch2, Ppof2, Prph2 <sup>*</sup> , Prss3 <sup>*</sup> , Rax <sup>*</sup> , Reep6, Rorb, Rrp1b, Samd11, Slc16a <sup>*</sup> , Slc17a <sup>*</sup> , Slc24 <sup>*</sup> , Slc38a <sup>*</sup> , Slc39 <sup>*</sup> , Slc4 <sup>*</sup> , Slc6a6, Tmem215, Tmem237 <sup>*</sup> , Trpm1 <sup>*</sup> |
| <b>Retina</b>                  | Abca4 <sup>*</sup> , Ankrd33 <sup>*</sup> , Ccdc <sup>*</sup> , Cdhrl, Chrna3a, Col <sup>*</sup> , Cryaa, Fscn2, Gucy2f, Irs1 <sup>*</sup> , Isl1, Kcnv2 <sup>*</sup> , Nr2e3, Nrl, Pde6a, Pde6b, Pde6h, Rdh8, Rho, Rpl, Rrh, Sag <sup>*</sup> , Sh2d <sup>*</sup> , Six3 <sup>*</sup> , Slc1a7, Tfap2 <sup>*</sup> , Vsx1, Vsx2  | Cryba <sup>*</sup> , Crybb2, Crygm <sup>*</sup> , Gabrr2, Glb1l2, Gnat1, Grm6, Isl2, Lgsn, Lim2, Mab2111, Opn1mw <sup>*</sup> , Opn4 <sup>*</sup> , Pax6 <sup>*</sup> , Prdm13, Prph <sup>*</sup> , Rcvrn <sup>*</sup> , Rgr, Rtbdn <sup>*</sup> , Samd7, Vax2   |

Enrichments in the pineal gland and retina relative to other tissues have been assessed by the determining the ratio of the normalized (+0.1 TPM) abundance of a transcript in each tissue relative to that in the mixed RNA sample (+0.1 TPM) to yield a relative expression value (rEx). Mixed RNA samples were made by mixing equal amounts of RNA from 6 to 20 tissues. The rEx values of the top 300 enriched transcript from the pineal gland and the top 300 from the retina were compared (pineal gland rEx/retina rEx) and transcripts that were > 10-fold were binned as pineal gland and those < 10 fold as retina; maximum levels were approximately 1000 for the pineal gland and 1/1000 for the retina. The remaining transcripts comprise the pineal gland and retina group. Zebrafish, mouse, rat and rhesus are included. The rat data are from the 24-hr time series experiment, the zebrafish data from the experiment with mixed tissue, and the rhesus and mouse data are from single experiments; the latter are from 129 mice. Data from all time points have been averaged and normalized (TPN + 0.1). The results below indicate whether a listed transcript is detected in all four or in only three species evaluated.

\* Asterisk, more than one homolog exists in some species; for example, Aanat<sup>\*</sup> represents Aanat in mouse, rat and rhesus in addition to Aanat1 and Aanat2 in zebrafish.