

Musical Genes: A Shiny Application to listen to the transcriptional symphony of parental care in pigeons

02 Feb 2021

Summary

In music, a “symphony” is an elaborate instrumental compositions written for an orchestra. In biology, the word symphony is used as a metaphor to describe the harmonious (or cacophonous) combination of elements that orchestrate cellular activities (e.g. 2020 Lemon Tijan). With RNA-sequencing, it is possible to quantify the expression of genes thought to regulate complex biological processes, but most RNA-seq workflows are designed to report results as figures and tables (rather than sound) and require bioinformatic expertise. To fill this gap, we built an R Shiny Application that allow interactive data exploration, visualization and sonification the transcriptomic symphony regulating parental care in pigeons. With a few clicks of a button, user can examine the raw data for genes of interest, listen to how the gene changes over time and relation to a few key reproductive hormones. This tools makes the research more accessible to scientists, peer-reviewers, and the general public.

Motivating dataset and data sonfication

This application uses transcriptomic data from Austin & Harris et al 2021 and hormone data from Austin et al 2021. These studies use a highly replicated experimental design to characterize the transcriptional activity of the hypothalamus, pituitary, and gonads axis at nine time points over the course of reproduction and parental care in male and female pairs as they transitioned from a sexually mature, non-reproductive state to a reproductive one. Nine pseudo-timepoints measured represent nest building, egg incubation, nesting care.

Data were imported into R and stored in duckdb for faster access. Additional information about genes, such as their description, ontology, and associated diseases were obtained from the gene ontology database (Genome alliance). Gene names for the human ortholog were imported from the HUGO database (ref). Data were joined, filtered, selected, mutated, and plotted with various R packages (). We used R shiny to build a web application with the dual ability to reproduce a figure from the associated manuscript and sonify the data (<https://raynamharris.shinyapps.io/musicalgenes/>).

To create the transcriptional symphony, we calculated the mean value of gene expression for each group and sonified those values using the R packages `sonify` and `tuneR` (ref). Additionally, the mean values were scaled from 0-6, rounded the data to an integer, and converted in letters from A to G so that the user can play a representation of the data on an instrument on their choosing, such as a piano. To explore the physiological symphony, we sonified the hormone concentration to see if it increases, decreases, or doesn't change as the gene of interest increases in expression. All the R code available under CC-BY at <https://github.com/raynamharris/musicalgenes>.

Significance

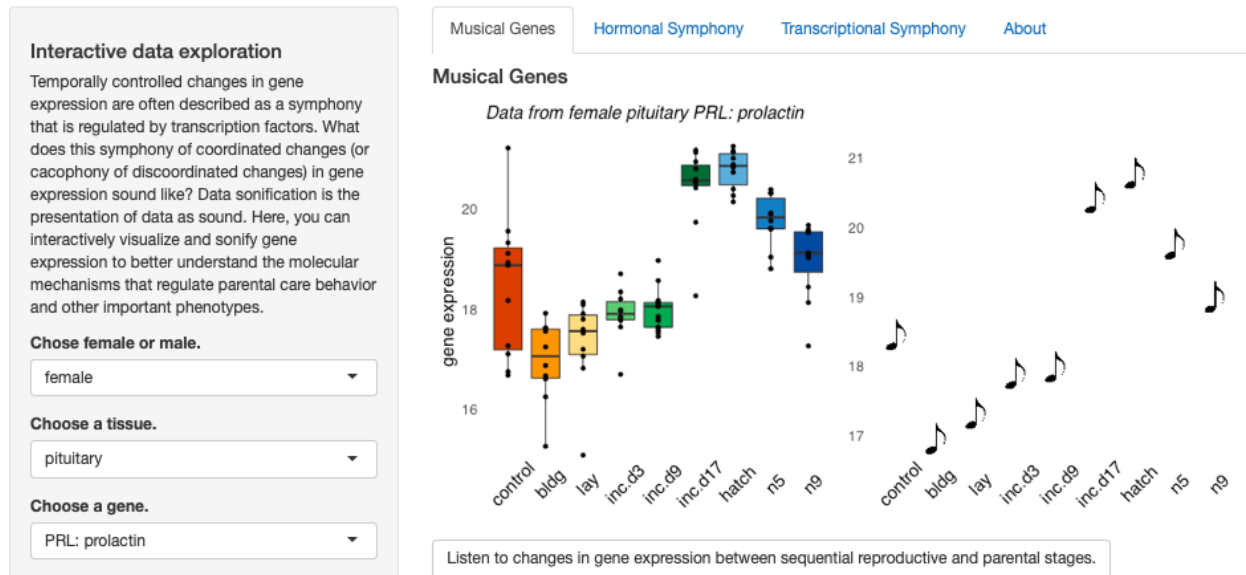


Fig. 1 Musical Genes

The default gene expression pattern shown by the shiny app is that of the prolactin gene (also referred to as PRL) in the female pituitary. These results show that prolactin increases significantly between incubation day 9 and incubation day 17.

Prolactin decreases significantly between non-reproductive control and nest building stages as well as between hatch and nestling care day 5. Prolactin decreases significantly when you remove offspring at incubation days 9, 17, and at hatch as well as when you extend the timing until hatch.

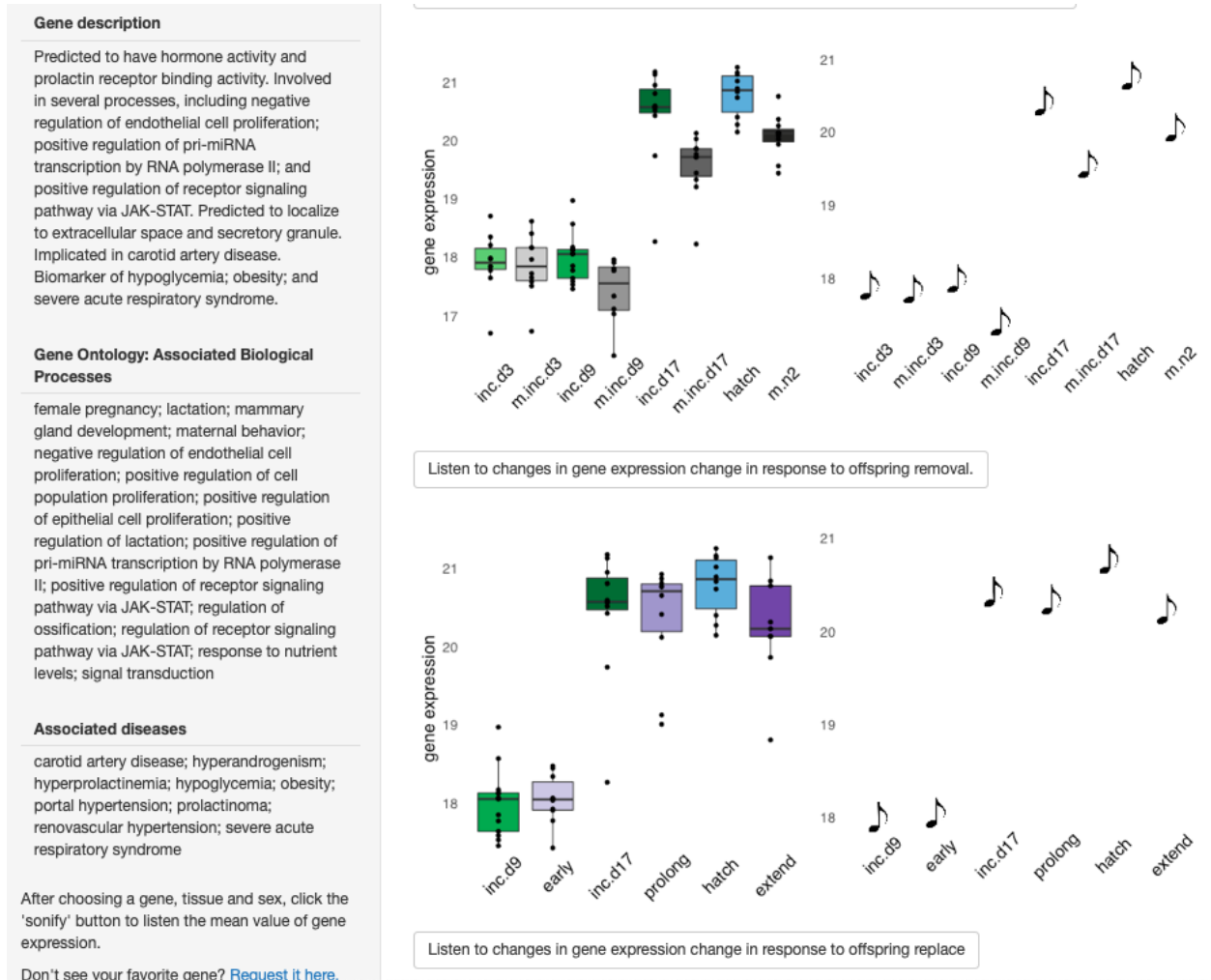


Fig. 2 Transcriptional cacophony?

The second tab seeks to integrate physiology and gene expression by showing correlations between circulating prolactin, sex steroids (estradiol and testosterone), corticosterone, and progesterone and the gene of interest.

Hormonal Symphony

Correlations between candidate genes and hormones

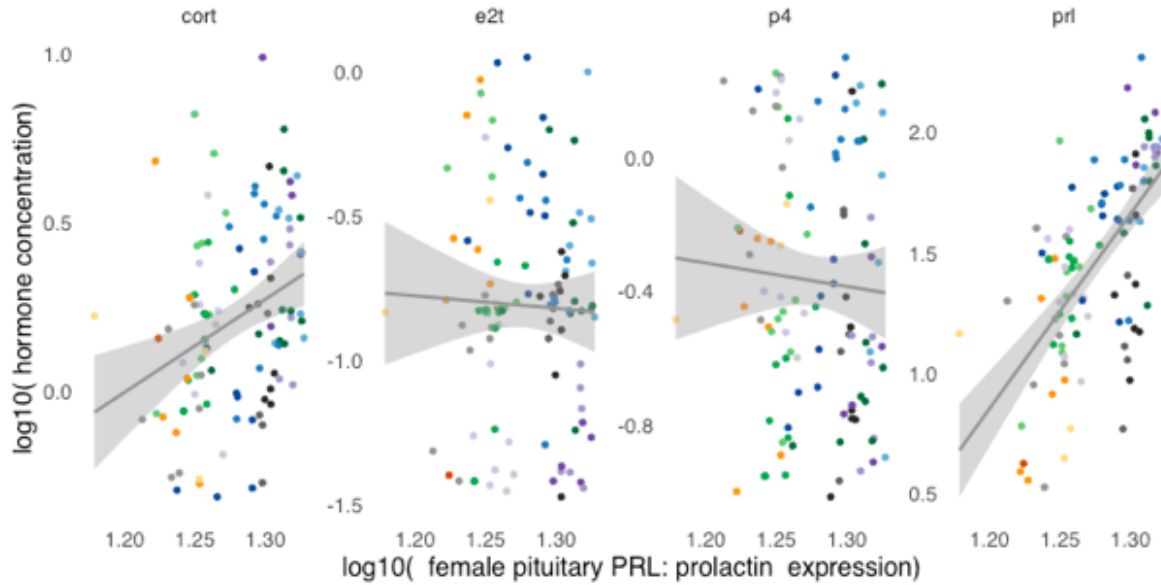


Fig. 3 Hormonal symphony

The third tab is about the future directions where we hope to tap the application to truly recreate the transcriptional and physiological symphony that regulates parental care. The final tab provides more information about the authors and the sources of funding for this project.

Transcriptional Symphony

In the future, we would like to be able sonify data for multiple genes or hormones simultaneously using sounds from instruments found in an orchestra.

gene_name	instrument	notes
ACOT12: acyl-CoA thioesterase 12	oboe	GCDBCAECEBFGDECB
APCDD1L: APC down-regulated 1 like	viola	CBFECFAGCBDFBGBEB
AVPR1B: arginine vasopressin receptor 1B	violin	BEDDEDEECEDAGDDF
BRCA2: BRCA2 DNA repair associated	trumpet	GBCBBCACDDDCDCCC
CKMT1A: creatine kinase, mitochondrial 1A	upright bass	EFDEFDGEDGDCFADF
COX2: NA	cello	GBCCBCBCACDCBCBC
CREBRF: CREB3 regulatory factor	clarinet	GDEAFBDFAGDAGEFF
CRHR2: corticotropin releasing hormone receptor 2	saxophone	BCDEDGDEDEBADFBE
DIO3: iodothyronine deiodinase 3	trombone	BDBDGBFBCGBADACC
NBN: nibrin	tuba	ECEFDCDDEGCADEBD
NPAS3: neuronal PAS domain protein 3	bassoon	GEGCFDGFACBCCCDC
NPY: neuropeptide Y	keyboard	GCDEBBDBABEBDCCF
PRLR: prolactin receptor	french horn	GABBBBCCCBABBBBB
TGFB1: transforming growth factor beta 1	piano	AGGFEEGFBDCDCCEE

Fig. 4 Transcriptional Symphony

Significance

This app allows for rapid data exploration associated with a research paper, which will allow peer-reviewers and readers without R expertise to explore the data.

We have created an Rstudio cloud environment that can sonify (turn into sound) the mean value of any given candidate gene and play it on the piano as a song as well as present resreent the data with statictics snd graphs as is typical of a scientific manuscript.

By providing the combination of a statistical summary, a visual representation, and a musical resprenstation of the data, one can look for both broad and specific patterns in the data. For instance, the results above for prolactin illustrate beautifully how some genes increase their expression in preparation for a biological even then fall once that even has past or if its timing is altered. This makes sense given what we know about prolactin... Should I say something about how co-authors of this paper are biased towards prolactin?? Like think these results to what we have written about its role in reproduction and parental care before (as in Oldfield, Harris, et al 2013; Austin and Word). Or say that Fararr et al in prep focuses solely on pituitary PRL in the hypothalamus and the gonas of male and females?

or

By simultaneously viewing the data, reading the statistical summary, and listening the the changes in mean expression, the user gains a deeper understand at the complexity of the variation in the data. The farther away the notes are on the scale, the greater the difference in expression. Smale changes in expression, as indicated by adjunct notes (e.g. F to G or E to D) are typically not significant, but large changes (e.g A to D or G to D) often represent significant increases or decreases in expression over time.

By creating an interactively webapplication, we facilitate the generation and exploration of the data and results for dozens of genes of interest. The avilability of this dataset will be of use to researchers trying to understand the role of candidate genes in regulating the fascinating biological transition into parenthood or the terrifying clinical transition into cancer. We turned gene expression data into music to make it possible to listen to transcriptional changes that are of interest to scientists and clinicians.

As stated in the introduction, we would like to bring to life the transcriptional symphony of parental care, but currently the tool Musical Genes only plays the sound of one gene at a time.

In the future, we would like to play the sounds of many genes at once by a digital or live orchestra to listen to coordinated or discordiated changes in gene expression.

One limiation of this application is that is is directly tied to this particular data. A future implementation of the tools could allow flexible data input. For instance, BioJupies is an online tool that rapidly creates interactive visualization of any RNA-seq data set in NCBI or a user uploaded data set (Torre et al). Our software applies to a single data set. The complexity of our experimental design with multiple variables and many combinations of comparison, scaling the data wrangling and visualization is not feasible. Regardless of size, future studies benefit from tools like BioJupies or Musical Genes or rapid data assessment. Despite these limitations, we still believe that this application presents a unique contribution to science by providing open access to software that makes data and results more accessible to diverse audiences.

References

Citations to entries in paper.bib should be in rMarkdown format.

If you want to cite a software repository URL (e.g. something on GitHub without a preferred citation) then you can do it with the example BibTeX entry below for (???).

For a quick reference, the following citation commands can be used: - @author:2001 -> “Author et al. (2001)”
- [@author:2001] -> “(Author et al., 2001)” - [@author1:2001; @author2:2001] -> “(Author1 et al., 2001; Author2 et al., 2002)”

Acknowledgements

We acknowledge contributions from x and y.

References