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 combination of elements that orchestrate cellular activities (Lemon and Tjian 2000). 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 (Reiter et al. 2021). To fill this gap, we built an R Shiny Application that allows interactive data exploration, visualization and sonification the transcriptomic symphony regulating parental care in pigeons. With a few clicks of a button, user can interactively 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, the general public, and to musicians who asipre to make data-inspire music. We hope this tool addressed some of the challenges that many researchers have face in turning big data into discovery (Stephens et al. 2015; Marder 2015).

Approach

This application uses transcriptomic data from Austin & Harris et al. (in prep) and hormone data from Austin et al. (in review). 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 Studio and stored in duckdb for faster access. Data were joined, filtered, selected, mutated, and plotted with various R packages (R Core Team 2019; Wickham et al. 2019; Wilke 2016). The mean value of gene expression was calculated for each group and sonified those values using the R packages `sonify` and `tuneR` (Ligges et al. 2018) (Fig 1). R Shiny (Chang et al. 2019) was used to build and deploy an interactive web application. (https://raynamharris.shinyapps.io/musicalgenes/). All the R code available under CC-BY at https://github.com/raynamharris/musicalgenes.

The sound works well when the shiny app is deployed locally, but it does not work on all browsers. To overcome this limiartion, we calculated and rounded the mean value of gene expression for each time point and and converted the value to the letters A to G so that the user can play a representation of the data on an instrument on their choosing, such as a piano (Fig 2). 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.

Next we sought to explore the phsiological symphony that results from the interplay between genes and hormones. Austin et al. 2021 in press, measured cirulating prolactin, sex steroids (estradiol and testosterone), corticosterone, and progesterone and found week correlations between prolactin and corticosterone. With this Shiny application, we can explore correlations between hormones and gene expression (Fig 3). To listen 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.

Because many users may not be familiar with many of the genes in this dataset, we important gene names for the human orthologs from the HUGO database (Braschi et al. 2019). Finally, we added additional information about genes, such as their description, ontology, and associated diseases were obtained from the gene ontology database (Agapite et al. 2020).

One limiation of this application is that is is directly tied to this particular data sature. 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, Lachmann, and Ma'ayan 2018). 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.

In summary, we have created an application that facilitates data exploration and hypothesis testing for scientist and clinicians that can also inspire artists to create music to better understand symphonies in a biolgoical context. This application is interactive, so it allows users to explore the raw data in a way that is not possible when reading a published manuscript saved as a pdf. 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 data and the biology and brings us one step closer to listening to the transcriptional symphony of parental care.

Acknowledgements

Figures

Fig 1.
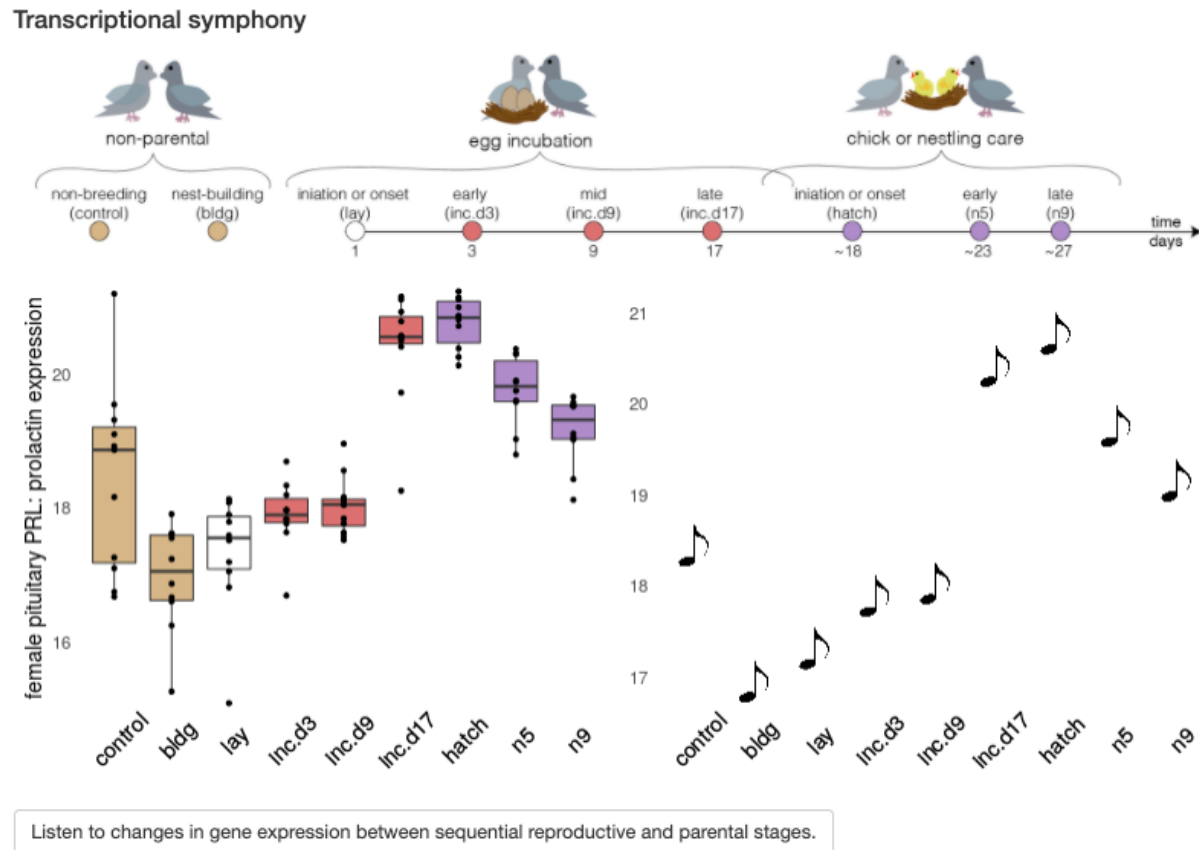
Fig 2.

gene	instrument	notes
AR: androgen receptor	violin	AGEEECCEF
AVP: arginine vasopressin	viola	GADAABABB
AVPR1A: arginine vasopressin receptor 1A	saxophone	BAEGEDABA
CRHBP: corticotropin releasing hormone binding protein	keyboard	GDFECABEC
ESR1: estrogen receptor 1	upright bass	AAGDCBCCA
FOS: Fos proto-oncogene, AP-1 transcription factor subunit	french horn	GACEDDBAB
GALR1: galanin receptor 1	clarinet	CECECEFGA
GNRHR: gonadotropin releasing hormone receptor	tuba	GGAEDDBCF
OXT: oxytocin/neurophysin I prepropeptide	cello	GABABABAA
PGR: progesterone receptor	trombone	EDACBAACG
PRL: prolactin	piano	CABBCFGED
PRLR: prolactin receptor	oboe	GAABBCBBA
TH: tyrosine hydroxylase	bassoon	BGBCBBABC
VIP: vasoactive intestinal peptide	trumpet	DCDGABCDE

Fig 3.

Listen to changes in hormone concentration as gene expression increases.
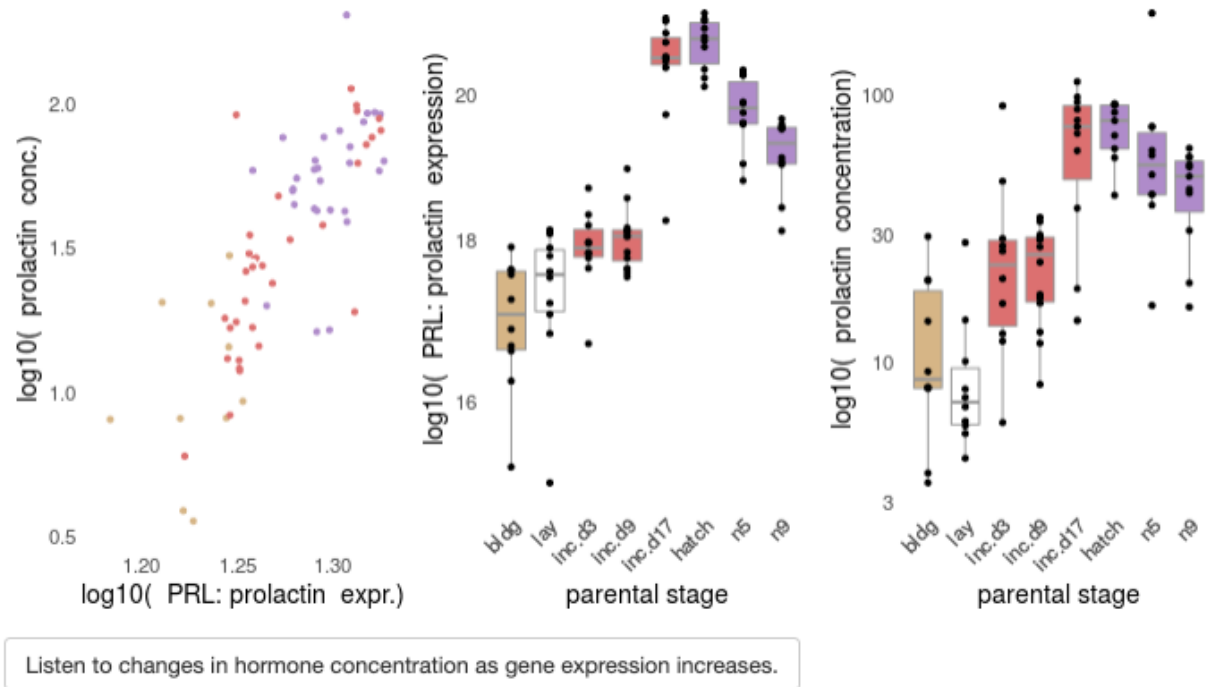
References

Agapite, Julie, Laurent Philippe Albou, Suzi Aleksander, Joanna Argasinska, Valerio Arnaboldi, Helen Attrill, Susan M. Bello, et al. 2020. "Alliance of Genome Resources Portal: Unified model organism research platform." *Nucleic Acids Research.* https://doi.org/10.1093/nar/gkz813.

Braschi, Bryony, Paul Denny, Kristian Gray, Tamsin Jones, Ruth Seal, Susan Tweedie, Bethan Yates, and Elspeth Bruford. 2019. "Genenames.org: The HGNC and VGNC resources in 2019." *Nucleic Acids Research.* https://doi.org/10.1093/nar/gky930.

Chang, Winston, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2019. *Shiny: Web Application Framework for R.* https://CRAN.R-project.org/package=shiny.

Lemon, Bryan, and Robert Tjian. 2000. "Orchestrated response: A symphony of transcription factors for gene control" 14 (20). https://doi.org/10.1101/gad.831000.

Ligges, Uwe, Sebastian Krey, Olaf Mersmann, and Sarah Schnackenberg. 2018. *tuneR: Analysis of Music and Speech.* https://CRAN.R-project.org/package=tuneR.

Marder, Eve. 2015. "Understanding Brains: Details, Intuition, and Big Data." *PLOS Biology* 13 (5): 1–6. https://doi.org/10.1371/journal.pbio.1002147.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reiter, Taylor, Phillip T Brooks, Luiz Irber, Shannon E K Joslin, Charles M Reid, Camille Scott, C Titus Brown, and N Tessa Pierce-Ward. 2021. "Streamlining data-intensive biology with workflow systems." *GigaScience* 10 (1). https://doi.org/10.1093/gigascience/giaa140.

Stephens, Zachary D., Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengxiang Zhai, Miles J. Efron, Ravishankar Iyer, et al. 2015. "Big Data: Astronomical or Genomical?" *PLOS Biology* 13 (7): e1002195. https://doi.org/10.1371/journal.pbio.1002195.

Torre, Denis, Alexander Lachmann, and Avi Ma'ayan. 2018. "BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud." *Cell Systems*. https://doi.org/10.1016/j.cels.2018.10.007.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the {tidyverse}." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wilke, Claus O. 2016. "cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'." https://cran.r-project.org/package=cowplot.