Musical Genes: Visualize and sonifiy gene expression in parenting pigeons

Rayna Harris and omany other[1] and Rebecca Calisi[1]

1 UC Davis

Summary

Background

Parental care is essential survival of many species, but it comes with many costs (). Caring for offspring means there is less time to attract new mates, and parents must balance their own needs with that of their offspring. How and why parents provide care for offspring has fascinated biologists for centuries (Champagne & Curley 1012). At the proximate level, a great deal is known about a handful of hormones and genes that promote or suppress parental care. Prolactin stimulates lactation, and vasopressin and oxytocin are known to increase affiliative behaviors, such as licking and grooming. Dopamine, serotonin, estrogen, and testosterone, on the other hand, are often associated with aggressive behaviors or avoidance. Much of what we know is from studies focused on a few genes in one tissue. However, no gene works alone, and gene expression regulation is tissue specific.

We have yet to conceptualize the "transcription symphony" of co-regulated gene expression that occurs throughout the reproductive axis to promote parental care. To test multiple variables and identify patterns of covariation, large scale studies are needed for sufficient power. Studying parental care in both sexes of mammalian species is challenging because only the females are able to use lactation to feed their young, and the physiological mechanisms regulating this process are hard to control for. The data from studies on model organisms are potentially useful to other scientists and clinicians. Genes are highly conserved across species. Gene networks are also conserved, but they may give rise to very different phenotypes in different species (aka, phenologs, Marcott lab publication). Therefore, it's possible that the gene networks that regulat parental care

Therefore but data accessibility is often a barrier to data reuse and thus the potential for biomedical discovery is diminished. Easy-to-use interfaces that allow users to interactively explore data in the cloud would speed discovery by eliminating the need to download data and install software as a precursor to data exploration. Sharing data and code alone, but using these notebook requires technical knowledge that many researchers and clinicians do not have. A challenge that researchers face is designing sequencing projects need to be large enough to have sufficient power to test hypotheses, and then making the data and results easily accessible to reduce barriers to translation research.

The biparental rock dove (Columba livia) is a useful model for studying transcriptional activity with the reproductive axis of both sexes (MacManes et al 2017, Calisi et al 2018, Lang et al 2020, Harris and Austin et al 2020). We built a shiny app so that the expression of 100 genes mentioned in this manuscript can be explored and evaluated in seconds. This research provides a deeper understanding of the interplay of genes, hormones, and parental care and has important implications for understanding molecular processes that underlie behavioral transitions. Our corresponding open source software and data has the potential to generate new questions, test existing hypotheses, and make science more accessible to diverse audiences.

Methods

Data accessibility

We aimed to make our data as findable, accessible, interoperable, and reusable (FAIR) as possible (Wilkinson et al 2016, Shew 2020). Color-blind friendly hex colors were chosen using `colorbrewer` (Brewer et al 2003). Mulit-panel figure are made with the aid of the graphing packages `cowplot` and `png` (Wilke, Urbanek). Data were sonified using the packages `sonify` and `tuneR` (Ligges et al 2018). Parts of the data behind figures 2-4 can explored interactively using https://raynamharris.shinyapps.io/musicalgenes/ (Fig.2 Suppl Fig. 2). The shiny app allows for explore of 100 genes in minutes, whereas reruning all the code from the workflow described above and illustrated in Fig. 2. Suppl. Fig. 3 would take days.

Data availability

RNA sequencing data are available through the European Nucleotide Archive (ENA) project ID PRJEB16136 and XXXXX. Code and data for reproducing the results described in the manuscript are available at https://github.com/macmanes-lab/DoveParentsRNAseq (upload to zendo and cite).

Significance

Listening to the transcriptional symphony with data sonification

Many neuroscientists use sound to evaluate the position of their electrodes inside cells. They are using sound to make inferences about the data. Sonification is the use of non-speech audio to convey information or perceptualize data. Molecular biology does not make use of such auditory cues and instead relies heavily on technicolor images, often with small font and color-blind insensitive palettes. We sought to improve the accessibility of our data by presenting the average gene expression values as notes on a scale (Fig. 2, Suppl Fig 2). Then, we used sonify and another R package, to turn the average value of gene expression for each group into tones (Suppl. File .wav). Here, we present one gene as computer generated tones, but one could imagine a pianist playing a few notes to understand how a handful of genes work together or a whole orchestra playing hundreds of genes working in symphony to regulate behavior.

Software for rapid and accessible data analysis.

So far, we have described numerous observations related to our a prior hypothesis and patterns that caught our eye during data exploration. This dataset that produces these insights has the potential to offer more questions and answers if it can be found and accessed by the right scientists and clinicians. To that end, we developed an online webpage using the R Shiny platform that allows users to recreate the boxplots shown in Figures 2-4 for all the genes mentioned in this manuscript (Fig 2. Suppl Fig 2). This covers candidate parental care genes from the literature and public databases, the top most differentially expressed genes, genes with strong co-expression patterns, and genes associated with breast, ovarian, testicular, and pituitary cancers. This allows readers to ask questions about their own candidate genes and get rapid answers, without having to dig into the workflow in depth (Fig 2. Suppl Fig 3).

Limitations

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.

Acknowledgements

References