

# Gene Symphony: Can Musification Capture Patterns in Expression Data ?

**Presented By :** Goutham B (BE20B010)  
**Guide:** Prof. Manikandan Narayanan

**Committee:** Prof. Karthik Raman, Prof. Nirav Bhatt

Indian Institute of Technology Madras



# Introduction

## What is Sonification?

- Translation of quantitative data into sound.
- Provides a new sensory channel for data analysis.

## What is Musification?

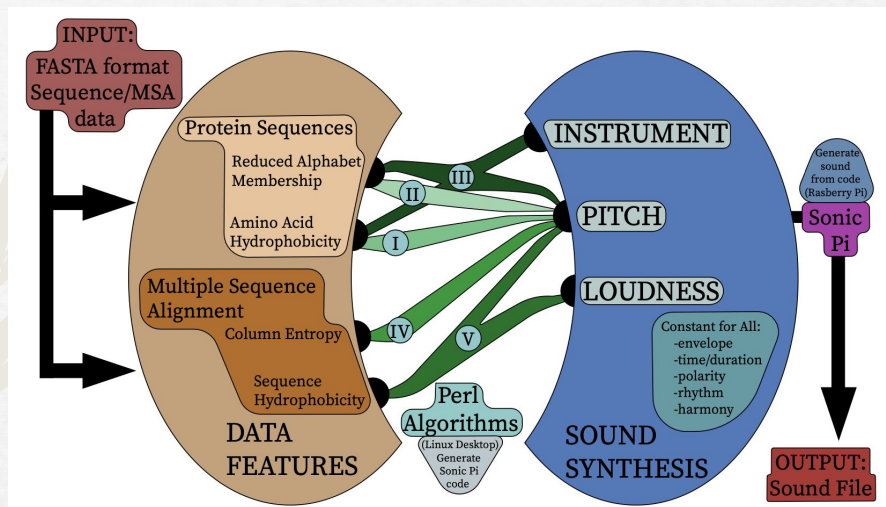
An extension of Sonification, which combines scientific data with musical composition for analytical and aesthetic purposes.

- Scientific research has relied heavily on visual data representations such as graphs and heatmaps.
- Visual representations can be limiting, particularly for those with visual impairments.
  - Blind: ~**3.44%** of total population have severe visual Impairments (based on a study in 2018\*).
  - Color blind: ~**4.5%** of total population.
- Challenging to interpret large volumes of information through traditional data presentation methods.
- Sonification, and by extension musification, bridges this gap by making data accessible through sound, offering a unique opportunity for alternative data analysis.

# Literature Survey

## Musification of DNA & Protein Sequences

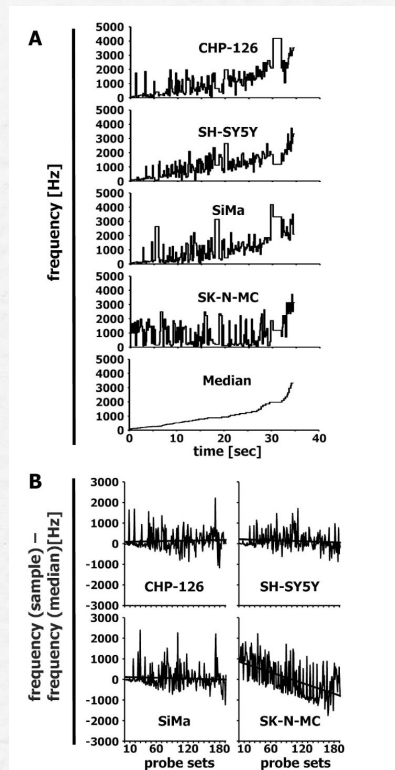
- Earliest Work: John Dunn's Presentation on "Musical Interpretations of DNA Data" in 1989.
- Recent work by Martin et al.(2021): Sonification techniques to analyse protein sequence data.



Martin et al.(2021) Overview of Sonification of Amino Acid Sequences & MSA.

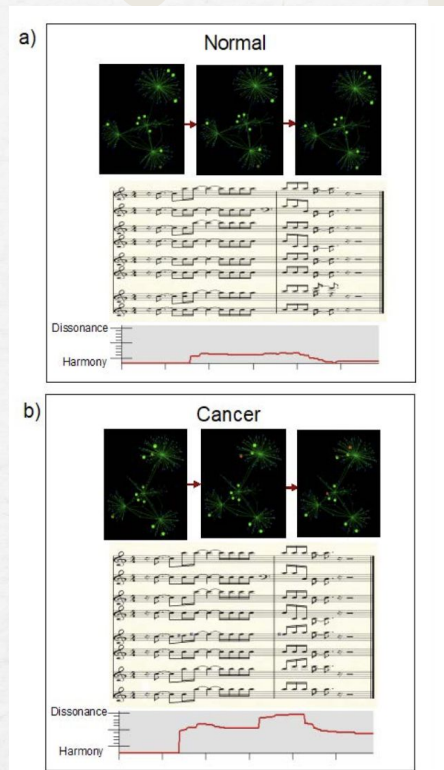
## Musification of Expression Data

### Static Expression Data



M.Staeger (2015) Frequency Plots.

### Time-Series Expression Data



Alterovitz et al. (2022) Musical Insights.



# Problem Statement

## MOTIVATION

- Sonification in biology is mostly limited to sequences & structure data.
- Previous works on expression data focus on simplifying expression values across the “**entire**” gene set into a “**single**” musical representation, which may lead to data loss and overlook localized expression dynamics.
- We bridge this gap by introducing a sonification mapping algorithm and music generation techniques that represent and analyze time-series gene expression through ‘**clusters**’, offering an alternative approach to interpreting gene expression heatmaps and dendrograms.

## AIM

***“To represent an expression vector heatmap using musical outputs as a novel way of analyzing time-series gene expression clusters”***

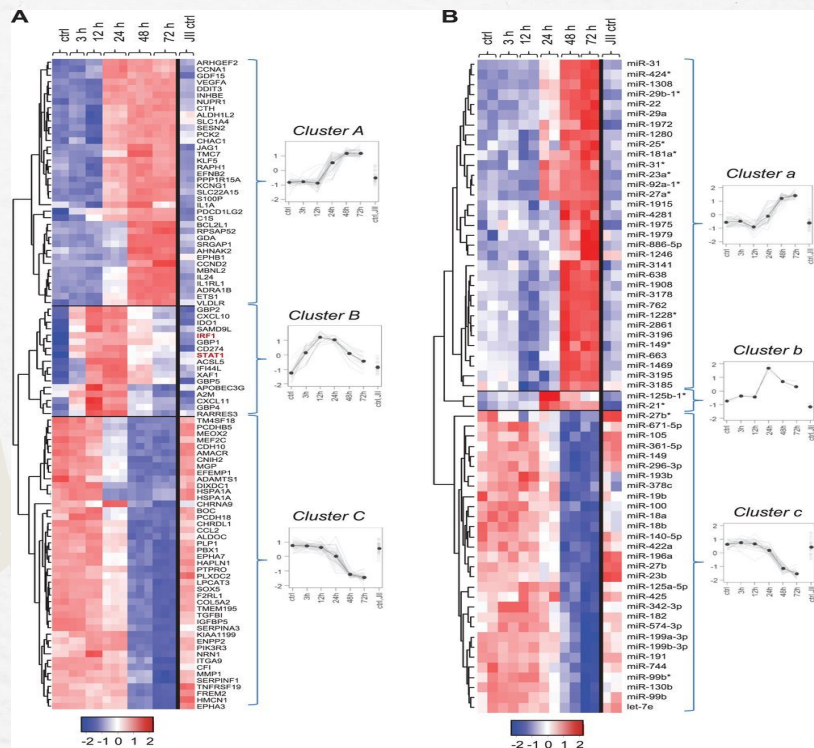
**Input:** A time-series gene expression vector of dimension  $\mathbf{G} \times \mathbf{T}$ , where  $\mathbf{G}$  is the number of genes and  $\mathbf{T}$  is the number of time points

**Output:** Multiple clusters (total  $\mathbf{C}$ ) of genes with similar expression patterns and their corresponding musical audio representations.

# Methodology

## 1. Preprocessing: Clustering

- Hierarchical clustering using Ward's method.
- Pairwise correlation for intra-cluster similarity.



## 2. Mapping Algorithm

$$N' = \left( \frac{V - V_{\min}}{V_{\max} - V_{\min}} \right) \times \alpha + \beta$$

Gao et al. (2022)

$$N = \begin{cases} \lfloor N' \rfloor & \text{if } \{N'\} < 0.5 \\ \lceil N' \rceil & \text{if } \{N'\} \geq 0.5 \end{cases}$$

$N$  is the MIDI note of the gene

$V$  is the expression value of the gene

$V_{\min}$  is the min expression value across all clusters

$V_{\max}$  is the max expression value across all clusters

$\alpha$  is set based on instrument (= 87 for piano)

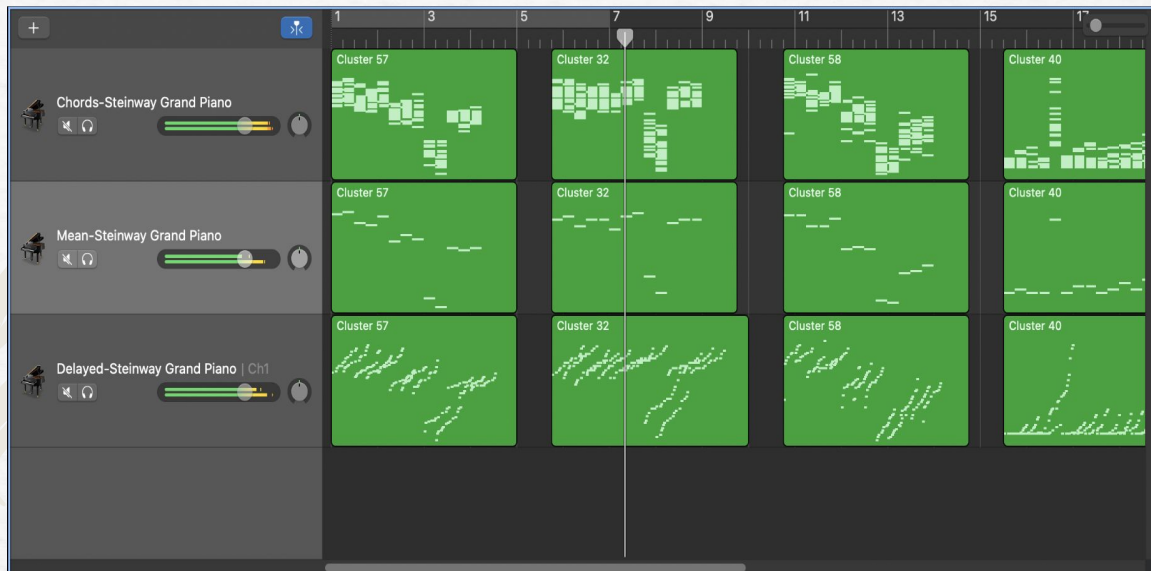
$\beta$  is a parameter which ensures notes are in the MIDI range (= 21 for piano)

MIDI: Musical Instrument Digital Interface

# Methodology

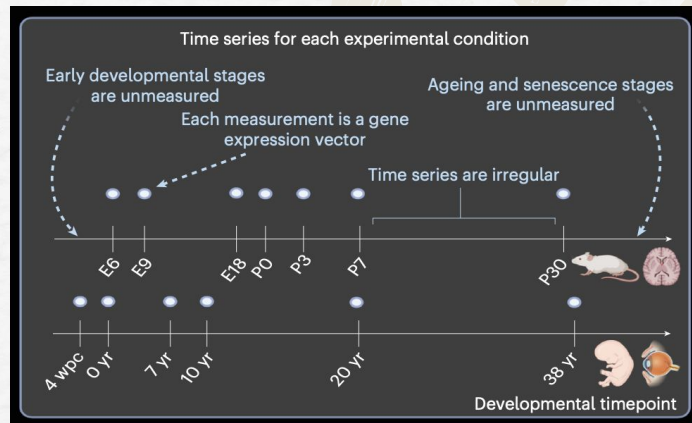
## 3. Generating Music & Variants

- **Chords:** Each time point forms a chord with all cluster genes represented as simultaneous notes.
- **Mean:** Each time point is represented by a single note based on the mean expression level.
- **Delayed:** Notes within a time point are played sequentially in ascending order, creating an “*arpeggio-like*” pattern.

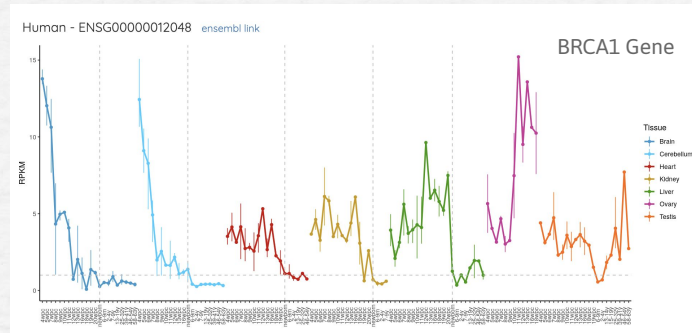


Utilizing the GarageBand Software for Music Variants Visualization

## 4. Data Used



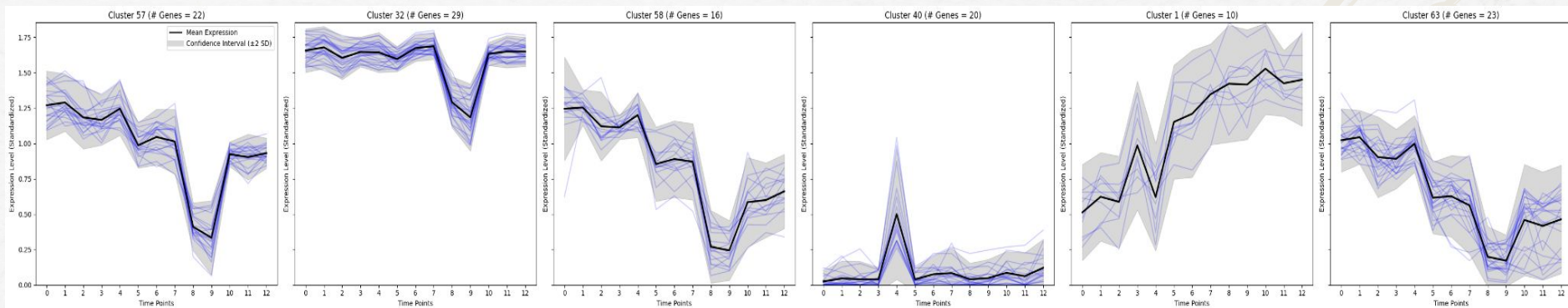
Woicik et al. (2023) Evo-Devo time points.



Kaessmann Lab, Expression Pattern of BRCA1 Gene in Human.

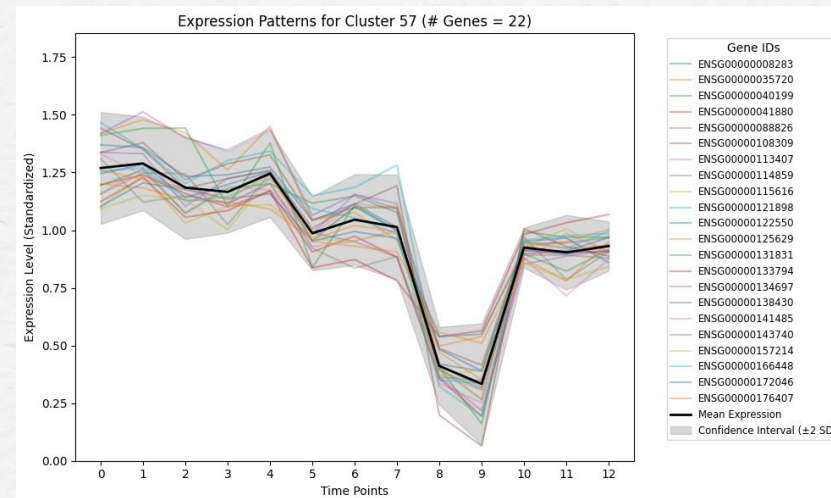


# Results



**Figure 1:** Expression Patterns for six top ranked clusters based on pairwise correlation.

TYPE/LABEL	“57”	“32”	“58”	“40”	“1”	“63”
CHORDS						
MEAN						
DELAYED						
COMBINED						



**Figure 2:** Expression Patterns for “57” containing 22 genes from the Human-Brain expression vector.

# Conclusion

- Each cluster's music captures expression trends, with high-pitched notes for highly expressed genes and low-pitched notes for lowly expressed ones.

Comparative Analysis of Different Studies on Sonification in Biology

Work	Type	Target Audience	Purpose	Data Used	Sonification Method
Martin <i>et al.</i> (2021)	Protein Sequences	Scientists & Visually Impaired Individuals	Alternate analysis for Protein Sequences & MSA	Multiple Protein Sequences (Insulin, Histones etc.)	Parameter Mapping Sonification (PMSon)
Franjou <i>et al.</i> (2019)	Protein Sequences & Structures	Any Individual	Aesthetics/Design New Proteins Using Music	Protein Sequences & Structure Data	Not mentioned
M. Staege (2015)	Static Gene Expression	Scientists & Visually Impaired Individuals	Differential Gene Expression	Expression Data for Neuroblastoma Cell Lines	Scaled Linear Mapping
Alterovitz <i>et al.</i> (2022)	Time-Series Gene Expression	Hospitals & Doctors	Patient Monitoring	Expression Data for Colon Cancer Cell	Pythagorean Tuning
<b>OURS (2024)</b>	<b>Time-Series Gene Expression</b>	<b>Scientists &amp; Visually Impaired Individuals</b>	<b>Alternate analysis for Clusters &amp; Heatmap/Aesthetics</b>	<b>Evolutionary Developmental Biology (Evo-Devo)</b>	<b>Normalized Linear Mapping</b>

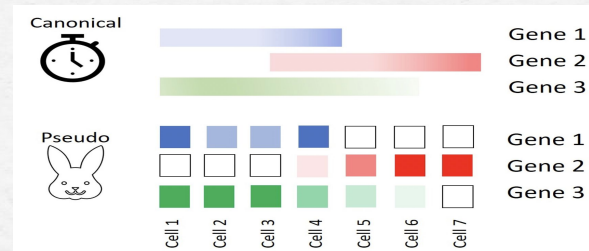


# Future Work

- **Evaluation Strategies:** Conduct group discussions or surveys where participants interpret graphs from musical samples to assess the effectiveness of music in conveying data patterns.
- **Enhanced Temporal Resolution:** Use tools like Sagittarius (Woicik et al., 2022) to extrapolate unmeasured gene expression data, improving the granularity of time-series musical representations.



- **Integration of Pseudotime Analysis:** Leverage pseudotime to map dynamic biological processes into music, reflecting temporal evolution and cellular state transitions.



Pseudotime Analysis.  
(Source: CellGS website\*)

- **Diverse Compositions with MusicLMs:** Employ Music Language Models to generate high-fidelity variations of musical representations, enriching perspectives and interpretability.

# Key References

1. **Martin, E.J., Meagher, T.R. & Barker, D.** Using sound to understand protein sequence data: new sonification algorithms for protein sequences and multiple sequence alignments. BMC Bioinformatics 22, 456 (2021). <https://doi.org/10.1186/s12859-021-04362-7>
2. **Staeger, M.** A short treatise concerning a musical approach for the interpretation of gene expression data. Sci Rep 5, 15281 (2015). <https://doi.org/10.1038/srep15281>
3. **Alterovitz G, Yuditskaya S** (2022) Musical Gene Expression: Abstracting High-Dimensional Gene Dynamics. J Biotechnol Biomater, 12: 280. DOI: 10.4172/2155-952X.1000280
4. **Petr V. Nazarov, Susanne E. Reinsbach, Arnaud Muller, Nathalie Nicot, Demetra Philippidou, Laurent Vallar, Stephanie Kreis,** Interplay of microRNAs, transcription factors and target genes: linking dynamic expression changes to function, Nucleic Acids Research, Volume 41, Issue 5, 1 March 2013, Pages 2817–2831, <https://doi.org/10.1093/nar/gks1471>
5. **Woicik, A., Zhang, M., Chan, J. et al.** Extrapolating heterogeneous time-series gene expression data using Sagittarius. Nat Mach Intell 5, 699–713 (2023). <https://doi.org/10.1038/s42256-023-00679-5>
6. **Franjou, S. L., Milazzo, M., Yu, C. H., & Buehler, M. J.** (2019). Sounds interesting: can sonification help us design new proteins? Expert Review of Proteomics, 16(11–12), 875–879. <https://doi.org/10.1080/14789450.2019.1697236>
7. **Gao, Z., Wang, H., Feng, G., & Lv, H.** (2022). Exploring sonification mapping strategies for spatial auditory guidance in immersive virtual environments. ACM Transactions on Applied Perception, 19(3), 1–21. <https://doi.org/10.1145/3528171>

# THANK YOU!