Journal of Biotechnology & Biomaterials

Short Communication   Open Access

Musical Gene Expression: Abstracting High-Dimensional Gene Dynamics

Gil Alterovitz[1,2,3]* and Sophia Yuditskaya[4]

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA
[2]Division of Health Sciences and Technology, Harvard Medical School and Massachusetts Institute of Technology, Boston, MA
[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA
[4]Cyber Analytics and Decision Systems Group at MIT Lincoln Laboratory, Cambridge, MA

Short Communication

Technologies for real-time monitoring of cellular protein abundance and gene expression are enabling personalized pharmacodynamic-based medicine, gene therapy, and patient evaluation – just as physicians have done in the past with various physiological signals [1-3]. Yet, the intrinsic complexities of real-time monitoring of multidimensional genomic and proteomic signals present new challenges. Novel methods are needed to fully take advantage of such technologies.

While current genomic and proteomic representations are visual, the demands of real-time monitoring with myriad inputs (whether it be in the lab or the operating room) necessitate alternative strategies. Humans have tens of thousands of genes and myriad proteins. It is impossible for a person to keep track of all of the gene expression and protein levels at any given time, unlike physiological factors – where there are only around twenty variables. This work describes an approach to representing the core information in genomic and proteomic data in a manner that can be easily presented and interpreted via music (see supplementary information and http://bcl.med.harvard.edu/proteomics/proj/Gn4D/).

Previous research has shed light on the utility of presenting static genomic sequences through music [4]. Others have worked on the sonification of genetic and protein structures, but not on comparisons of their expressions or interactions [5]. This work focuses on extracting key information from the dynamics of genes or proteins levels – by looking at the gene expression/protein abundance, over time, through sound. While a human's sequence is generally static, gene expression and protein abundance can change across time – and across different tissues. Being able to use such information and effectively translate gene/protein levels into actionable sound, to guide decision making in real-time, could have tremendous impact on patient monitoring, biological investigation, and many other domains that necessitate real-time analysis.

The temporal quality of music, coupled with alerts which can be triggered by inharmonious music, alarms, and other mechanisms, provides a means to present status information – without requiring the full attention of the user that a visual monitor would. In addition, sound can also reveal aspects of multidimensional data that may be hidden in visual representations [6].

In addition to high multidimensionality of the data, another challenge is the absence of numerical thresholds and standards. In clinical medicine, a plethora of such criteria exist and has helped to standardize clinical measurements. For example, a heartbeat of greater than 100 beats per minute for adults is tachycardia, while it would be considered bradycardia for infants. Yet, baseline genomic and proteomic signals are becoming increasingly available from normal / control patients – and our method seeks to use this information. Here we present a method for comparing different states through sound, referred to herein as comparative sonification, for real-time monitoring of patient health and other mission-critical applications. Sonification of genetic and proteomic expression data provides an alternative medium (sound) for its presentation, which can facilitate analysis and discovery. Sound is a powerful means of transmitting information, both alone and coupled with visual cues, because it can significantly increase the bandwidth of human/computer interfaces.

In our approach, we first create an abstract representation of gene expression or protein abundance across time by extracting out the fundamental principal components of expression patterns (see supplementary information) [7]. Originally, each gene represents one dimension. So, thousands genes have thousands of dimensions. It is possible to represent many genes in fewer dimensions by finding the fundamental principal components. This was applied to a colon cancer study dataset, where we abstracted over three thousand genes to four dimensions that are linear combinations of the original set of genes [8]. We then presented this information in an inherently time-based medium, namely music. Control samples were normalized such that their musical notes mapped to harmonious intervals in the musical scale – and were then assigned instruments to each of the chief components. We determined the frequencies for the experimental test data set, which can be normal or perturbed, by comparing the principal component analysis based-adjusted expression values to their analogs (same gene, same time) in the control data set. Using Pythagorean tuning mathematics, the control notes were then normalized to harmonious intervals [9]. The nth note at a given time is the projection of all genes in a particular dimension n, for a total of n different notes. So, when playing three or more notes at the same time, we get chords. When played across time, we get music.

This approach allowed us to map relative changes in expression to the musical scale in a principled manner. The result is that the experimental data mapped to the same notes as the control data only if it fluctuates in the same manner as the control. Figure 1 shows the results of when a colon cancer and normal experimental datasets (GEO: GDS16138) were compared to control data. We found that normal samples, which typically fluctuate in a similar manner to the control, resulted in harmonious music. In addition, we found that the cancer samples sounded inharmonious across time and experimental conditions based on both auditory and quantitative metrics (Music Graph, Composer Tools, Annapolis, MD) [10]. Interestingly, the dynamics of cancer samples resulted in more inharmonious sound

---

*Corresponding author: Gil Alterovitz, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA, E-mail: gil_alterovitz@hms.harvard.edu

---

Citation: Alterovitz G, Yuditskaya S (2022) Musical Gene Expression: Abstracting High-Dimensional Gene Dynamics. J Biotechnol Biomater, 12: 280.

Page 2 of 2
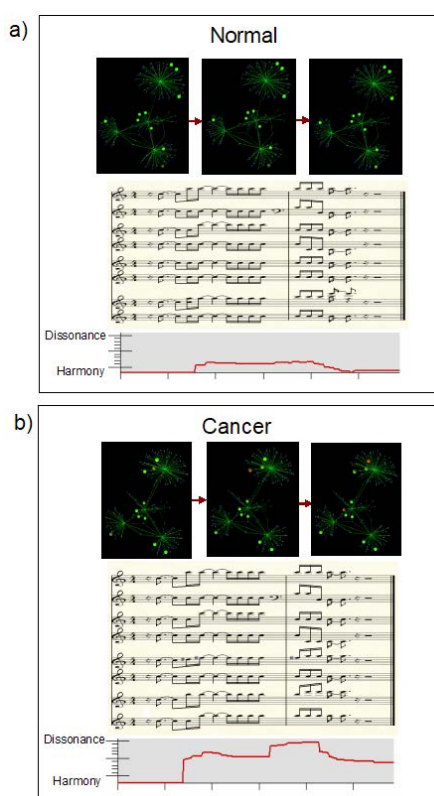
Figure 1: a) Time course for a) normal and b) p53 null mutant colon cancer cells in response to inflammatory stress conditions [7]. In each panel, the top segment shows the gene nodes colored based on expression levels. Links are based on the underlying protein interactions. Using method described in this work, gene expression's principal components are mapped to the music scale. The result is a harmonic musical sequence shown in a), versus the discordant one as depicted in b). Here, the staffs are played by harpsichords, recorders, flutes, and oboes in the Baroque style.

across time as the state worsened – suggesting that the transitions led to loss of control as variability increased for neighbors within the underlying protein network (Figure 1).

We have also used the auditory and visual display to provide an interactive way to explore microarray expression analysis and protein interaction networks for physicians and scientists. A number of studies have established that learning via multiple modalities (e.g. auditory and kinesthetic in addition to visual [6]) improves knowledge retention. We have found that students and research trainees were enthusiastic about learning through this interactive approach.

With new real-time genomic and proteomics monitoring technologies, novel methods are needed to fully utilize the benefits afforded by such techniques. The wealth of such molecular measurements brings with it new challenges to patient monitoring. Through the approach described here, this new information can be analyzed in a way that does not overwhelm already inundated physicians and biomedical researchers.

Acknowledgement

References

1. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Sci 311: 1600-1603.

2. Beliën A, De Schepper S, Floren W, Janssens B, Mariën A, et al. (2006) Real-time gene expression analysis in human xenografts for evaluation of histone deacetylase inhibitors. Mol Cancer Ther 5: 2317-2323.

3. King KR, Wang S, Irimia D, Jayaraman A, Toner M, et al. (2007) A high-throughput microfluidic real-time gene expression living cell array. Lab Chip 7: 77-85.

4. Takahashi R, Miller JH (2007) Conversion of amino-acid sequence in proteins to classical music: search for auditory patterns. Genome Biol 8: 405.

5. Mössinger J (2005) Science in culture: The music of life. Nature 435.

6. Conway CM, Christiansen MH (2005) Sequential Learning by Touch, Vision, and Audition. J Exp Psychol Learn Mem Cogn 31: 24-39.

7. Joliffe I (2002) Principal Component Analysis Springer.

8. Staib F, Robles AI, Varticovski L, Wang XW, Zeeberg BR, et al. (2005) The p53 tumor suppressor network is a key responder to microenvironmental components of chronic inflammatory stress. Cancer Res 65: 10255-10264.

9. Leech-Wilkinson D (1997) Companion to Medieval and Renaissance Music. Oxford University Press, UK.

10. Forte A (1973) The Structure of Atonal Music. Yale University Press, New Haven, London, ix: 224.