Glass Half Full or Empty: Illuminating the Human Transcriptome

Inspiring Underserved High School Students with an Experiential Biology Laboratory Research Experience

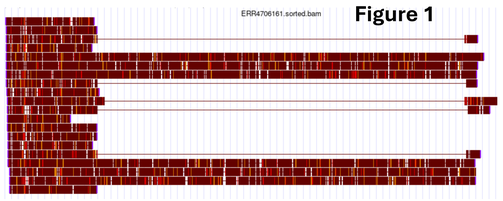
Theodore M. Nelson[1,2,3,4,5], Dennis J. Zhang[1], Sophia Sorid[1], Timothy Chang[1], Neil Bajaj[6], Joseph E. Viola[7], Lauren N. Heckelman[7], Robert E. Pollack[1]

[1] Department of Biological Sciences, Columbia University, New York, NY 10027, USA; [2] Department of Microbiology & Immunology, Vagelos College of Physicians & Surgeons, Columbia University Irving Medical Center, New York, NY 10032, USA; [3] Department of Systems Biology, Vagelos College of Physicians & Surgeons, Columbia University Irving Medical Center, New York, NY 10032, USA; [4] Department of Physiology and Biophysics, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA; [5] BioAstra, Inc, New York, NY, USA; [6] Department of Chemistry, Columbia University, New York, NY 10027, USA; [7] Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA

Protein diversity found in higher eukaryotic organisms can be attributed to alternative splicing, whereby RNA transcripts are processed into different isoforms. In humans, up to 95% of all multi-exon genes are alternatively spliced and ~15% of inherited diseases are associated with this process.[1]

Class #1:
The RNA World
LRS Analysis

Figure 1



The Long-Read Analysis Pipeline for Transcriptomics[2] (L-RAPiT) was employed on six published human HEK 293T LRS samples: SRR14228534, SRR12389274, SRR10832427, ERR4972057, ERR4837073, ERR4706161. Within L-RAPiT, StringTie & gffcompare were utilized to identify novel transcripts. Novel transcripts replicated in all six samples were selected for wet-lab verification. Sample alignments were visualized in the hg38 UCSC genome browser (Figure 1, above). The exemplar spliced transcript is called "TCONS_00013611" and is located on chr10 between chromosomal locations 73,742,581-73,744,844.

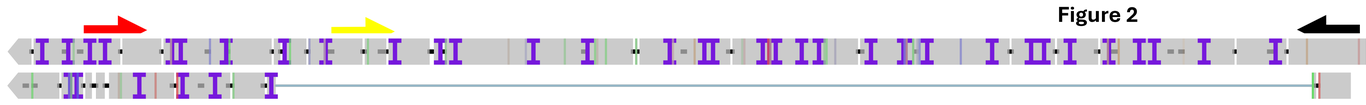
Class #2:
Primer Design

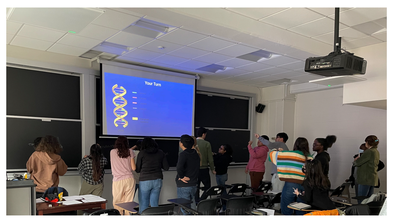
Class #3:
Safety Training

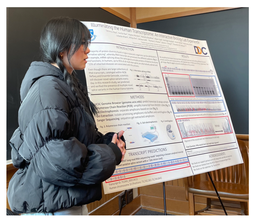
Class #4:
Polymerase Chain Reaction

Students analyze the data, noting via the UCSC Browser that the transcript overlaps with the unspliced "ENSG00000279088", present within the Ensembl human GENCODE V44 reference transcriptome. The students design primers to check for the presence of both the known unspliced and novel predicted unspliced variant (Figure 2, below).

Figure 2







Class #5:
Gel Electrophoresis

Class #6:
Gel Extraction

Class #7:
Sanger Sequencing

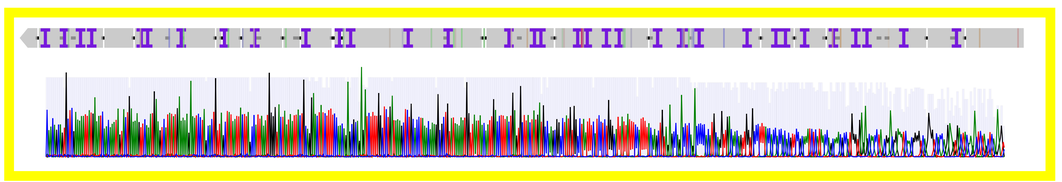
Class #8:
Poster Presentation

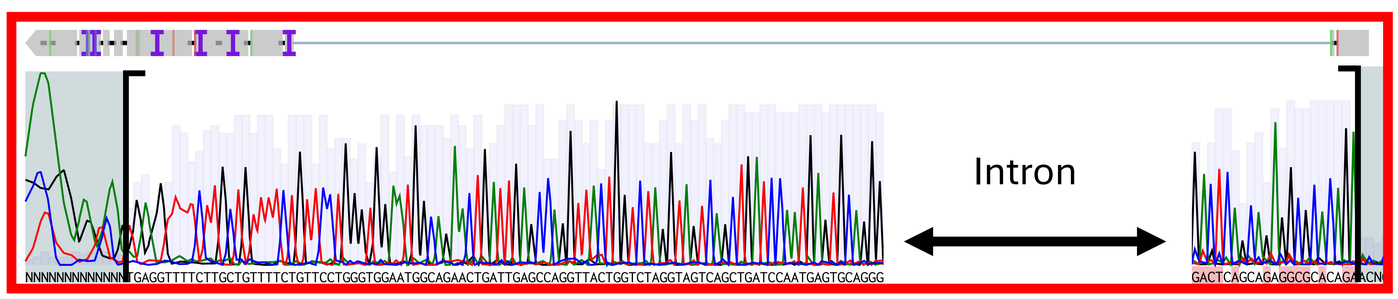
Results: The course is currently in its fourth iteration, having fully served 42 low-income students from the greater New York City community. The students have verified ten novel transcript variants, contributing to a more complete version of the human transcriptome. They become researchers working within a reproducible scientific tradition, recapitulating the results of their classmates across multiple independent replicates. The course includes a volunteer program, allowing to date 35 undergraduates to learn with and mentor participants. Each class is associated with a check-out quiz, which, along with post-class surveys, exemplify retainment of the material and an increase in confidence as it relates to the practice of scientific research.

As one two-time student wrote, "I chose to take this class course the first time around because I've always been interested in biology. The high school I go to does not give me the advantage to take biology related courses; however, DDC did!! This class has impacted my future planning. As a senior in high school, I am still unsure about what I would like to major in. Though, after learning more about biology, I've considered majoring in biology!!"

Future Directions: We are expanding the availability of course resources via YouTube on MakeTheBrainHappy – Scientific Exploration, to make the course findable, accessible, and reproducible for instructors.

Figure 3: Sanger sequencing alignments for both known (yellow) and novel transcript TCONS_00013611 (red), signifying wet-lab verification of transcript existence.





1. Jiang, W.; Chen, L. Alternative Splicing: Human Disease and Quantitative Analysis from High-Throughput Sequencing. *Computational and Structural Biotechnology Journal* **2021**, *19*, 183–195, doi:10.1016/j.csbj.2020.12.009.

2. Nelson, T.M.; Ghosh, S.; Postler, T.S. L-RAPiT: A Cloud-Based Computing Pipeline for the Analysis of Long-Read RNA Sequencing Data. *IJMS* **2022**, *23*, 15851, doi:10.3390/ijms232415851.

Department of Biomedical Engineering COLUMBIA | ENGINEERING

ASBMB American Society for Biochemistry and Molecular Biology

Science And Subjectivity

THE ROGER LEHECKA Double Discovery Center at Columbia University

Stephen's Family Foundation

SBI