

An undergraduate bioinformatics curriculum that teaches eukaryotic gene structure

Meg M. Laakso^{1*}, Leocadia V. Paliulis², Paula Croonquist³, Brianna Derr², Elena Gracheva⁴, Charles Hauser⁵, Carina Howell⁶, Christopher J. Jones⁷, Jacob D. Kagey⁸, Jennifer Kennell⁹, S. Catherine Silver Key¹⁰, Hemlata Mistry¹¹, Srebrenka Robic¹², Jamie Sanford¹³, Marisol Santisteban¹⁴, Chiyedza Small¹⁵, Rebecca Spokony¹⁶, Joyce Stamm¹⁷, Melanie Van Stry¹⁸, Wilson Leung⁴, and Sarah C.R. Elgin⁴

¹Eastern University

⁷Moravian College

¹³Ohio Northern University

²Bucknell University

⁸University of Detroit Mercy

¹⁴The University of North Carolina at Pembroke

³Anoka-Ramsey Community College

⁹Vassar College

¹⁵Medgar Evers College, The City University of New York

⁴Washington University in St. Louis

¹⁰North Carolina Central University

¹⁶Baruch College, The City University of New York

⁵St. Edward's University

¹¹Widener University

¹⁷University of Evansville

⁶Lock Haven University

¹²Agnes Scott College

¹⁸Lane College

Abstract

Gene structure, transcription, translation, and alternative splicing are challenging concepts for many undergraduates studying biology. These topics are typically covered in a traditional lecture environment, but students often fail to master and retain these concepts. To address this problem we have designed a series of six Modules that employ an active learning approach using a bioinformatics tool, the genome browser, to help students understand eukaryotic gene structure and functionality. Students learn how to use a mirror site of the UCSC Genome Browser created by the Genomics Education Partnership while completing the Modules, which focus on gene structure, transcription, splicing, translation, and alternative splicing. The Modules are supplemented with short videos that illustrate key functionalities of the genome browser and fundamental concepts in processing transcripts. These materials have been used successfully to teach gene structure in many different settings, from community colleges to 4-year colleges and universities, encompassing advanced high school students to college seniors. Instructors can easily customize the Modules and/or select a subset for their curriculum. The Modules have helped our students learn about eukaryotic gene structure and expression, simultaneously acquiring skills in the use of a genome browser, and have prepared them to pursue genome annotation projects as independent research.

Citation: Laakso, M.M., Paliulis, L.V., Croonquist, P., Derr, B., Gracheva, E., Hauser, C., Howell, C., Jones, C.J., Kagey, J.D., Kennell, J., Silver Key, S.C., Mistry, H., Robic, S., Sanford, J., Santisteban, M., Small, C., Spokony, R., Stamm, J., Van Stry, M., Leung, W., Elgin, S.C.R. 2017. An undergraduate bioinformatics curriculum that teaches eukaryotic gene structure. *CourseSource*. <https://doi.org/10.24918/cs.2017.13>

Editor: Mark Pauly, University of Nebraska-Omaha, Omaha NE

Received: 12/01/2016; **Accepted:** 03/17/2017; **Published:** 06/26/2017

Copyright: © 2017 Laakso, Paliulis, Croonquist, Derr, Gracheva, Hauser, Howell, Jones, Kagey, Kennell, Silver Key, Mistry, Robic, Sanford, Santisteban, Small, Spokony, Stamm, Van Stry, Leung, and Elgin. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited. No permission is required from the authors or the publishers for noncommercial use of the published materials. The authors affirm that they either own the copyright to or have received written permission to use the text, figures, tables, artwork, abstract, summaries and supporting materials.

Conflict of Interest and Funding Statement: None of the authors have a financial, personal, or professional conflict of interest related to this work.

Supporting Materials: S1. Module 1: What is a Gene?, S2. Module 2: Transcription Part I, S3. Module 3: Transcription Part II, S4. Module 4: Splicing, S5. Module 5: Translation, S6. Module 6: Alternative Splicing, S7. Browser video, S8. Genes and isoforms video, S9. RNA-Seq and TopHat video, S10. Short Match video, S11. Splicing and Phase video, S12. Tracks video, S13. Module 1: What is a Gene? Answer Sheet, S14. Module 2: Transcription Part I Answer Sheet, S15. Module 3: Transcription Part II Answer Sheet, S16. Module 4: Splicing Answer Sheet, S17. Module 5: Translation Answer Sheet, S18. Module 6: Alternative Splicing Answer Sheet, S19. Glossary of terms, and S20. Sex determination mini-lecture.

*Correspondence to: 1300 Eagle Road, St. Davids, PA, 19087

Email: mlaakso@eastern.edu

Learning Goal(s)

Students will understand the structure of a eukaryotic gene, and how sequences within a gene impact the gene product. Students will understand the relationships among DNA, pre-mRNA, mRNA, and protein. Students will also learn how to use a genome browser, and to interpret the evidence tracks produced by multiple bioinformatics tools (e.g., genome sequence, RNA-Seq data) to gain an in-depth understanding of eukaryotic genes.

Learning Objective(s)

Module 1

- Demonstrate basic skills in using the UCSC Genome Browser to navigate to a genomic region and to control the display settings for different evidence tracks.
- Explain the relationships among DNA, pre-mRNA, mRNA, and protein.

Module 2

- Describe how a primary transcript (pre-mRNA) can be synthesized using a DNA molecule as the template.
- Explain the importance of the 5' and 3' regions of the gene for initiation and termination of transcription by RNA polymerase II.
- Identify the beginning and the end of a transcript using the capabilities of the genome browser.

Module 3

- Explain how the primary transcript generated by RNA polymerase II is processed to become a mature mRNA, using the sequence signals identified in Module 2.
- Use the genome browser to analyze the relationships among:
 - pre-mRNA
 - 5' capping
 - 3' polyadenylation
 - splicing
 - mRNA

Module 4

- Identify splice donor and acceptor sites that are best supported by RNA-Seq data and TopHat splice junction predictions.
- Utilize the canonical splice donor and splice acceptor sequences to identify intron-exon boundaries.

Module 5

- Determine the codons for specific amino acids and identify reading frames by examining the Base Position track in the genome browser.
- Assemble exons to maintain the open reading frame (ORF) for a given gene.
- Define the phases of the splice donor and acceptor sites and describe how they impact the maintenance of the ORF.
- Identify the start and stop codons of an assembled ORF.

Module 6

- Demonstrate how alternative splicing of a gene can lead to different mRNAs.
- Show how alternative splicing can lead to the production of different polypeptides and result in drastic changes in phenotype.

INTRODUCTION

In 1984, Benjamin Bloom coined the phrase “the 2 sigma problem” to describe the observation that the average student who received individual tutoring using “Mastery Learning” techniques showed a two standard deviation advantage compared to a control group of students (i.e., they perform better than 98% of the students in the control group) (1). Hence an important challenge in teaching is to identify strategies that could be used in a large classroom that would result in the same learning gains as those achieved with one-on-one instruction.

In science, technology, engineering and mathematics (STEM) fields, one-on-one tutoring typically occurs within the undergraduate or graduate research setting. Many published studies have confirmed the benefits of this approach for training future scientists (2,3): research experiences often

increase interest in biology, and can help a student develop an identity and sense of confidence as a scientist. Unfortunately, one-on-one mentoring requires substantial financial resources, lab space, and time commitments from mentors, so this is not a scalable approach to training students. In fact, in many educational institutions, this level of individual attention is not possible at all.

During the past 30 years, a multitude of practices have been proposed as pathways to improving STEM education in the classroom setting (4). Active learning and collaborative/small group learning have been shown to have significant positive effects on learners compared to the traditional lecture-based method of teaching (5,6). Since the early 2000's, evidence has been accumulating that beginning students and underprepared students can close the achievement gap with their more advanced and better prepared peers when the classroom learning environment is structured in this way (7).

The lecture model for teaching has lost favor in recent years because it results in lower learning gains compared to other models (8). This might be partially related to cultural shifts; undergraduates report spending eight or more hours per day on their phones, for communication, entertainment, and social networking (9), and view the world as a complex network of relationships. Using computer-based curricula leverages these changing social norms and student familiarity with technology to create a learner-centered environment (10). Setting up a learning environment where students work in a collaborative group also promotes the social aspect of learning (11,12).

Nonetheless, it remains a challenge for faculty to adopt active learning strategies because of the substantial time and resources needed to develop the curricula. These valuable resources are often in short supply (13,14). However, in the life sciences, an increasing number of resources are available for faculty who want to incorporate genomics and bioinformatics components into their courses, some of which use an active approach (15, 16 and references therein). This includes a Bioinformatics Learning Framework for the integration of bioinformatics into new or existing courses (17), as well as lesson plans and curricula that could potentially be used by faculty or their students. These range from freely available Massive Open Online Courses (MOOCs) and bioinformatics tutorials to subscription services and online university courses (e.g., MOOCs through University of California, San Diego at <https://www.coursera.org/ucsd>; Open Helix tutorials at <http://openhelix.com/freeTutorials.cgi>; Coursera subscription courses at <https://www.coursera.org/>; and Harvard Extension courses at <https://www.extension.harvard.edu/>, respectively). The use of prepared curricula can mitigate some of the challenges, but the learning curve for the instructor often remains steep and little time may be available to master the new skills that are necessary to customize and use the prepared curricula effectively.

The six Modules presented here (Supporting files S1-S6) allow the instructor to complete class preparation and mastery of the biological concepts and bioinformatics skills required in a manageable amount of time. Genomics Education Partnership (GEP) faculty who have used the curriculum report spending 30-60 minutes in preparation for each Module, a significant time savings compared to the many hours needed to develop and write one's own guided curriculum. The biological concepts covered in these Modules will already be familiar to biology faculty, as they include foundational concepts in genetics and molecular biology such as gene structure, transcription and translation, and alternative splicing. Although it would be beneficial for the faculty to have some prior experiences in bioinformatics, expertise or research experience in the field is not a prerequisite for the successful use of these Modules. Instructors do not need to attend workshops or training courses prior to using these Modules, since they are self-guided. Six short videos are also associated with the Modules (Supporting files S7-S12). These videos illustrate difficult concepts and provide background information on the genome browser and its associated functionalities. Students also may view the videos once or multiple times to introduce and reinforce new concepts and skills.

An additional benefit of this bioinformatics curriculum is the ease with which this active learning approach can be adopted by undergraduate faculty in the classroom. The lessons we propose here require few resources: a computer, an Internet connection, and the guidance of the instructor. We have found that most of our students have a personal laptop

and are willing to bring it to class. Alternatively, students can use computers in a computer lab or check out a laptop to work on these Modules during the class sessions.

The advantage of using these Modules, beyond their "hands on" approach, lies in the facility with which the browser can visually display a eukaryotic gene. A student can very easily shift from a "whole gene" view, as when examining the pattern of exons and introns, to a "zoomed in" view, allowing quick identification of the methionine codon at the beginning of the coding sequence. The simultaneous translation and presentation of the three reading frames clarifies the meaning of reading frames, and the marking of stop codons in red allows the viewer to pick out open reading frames with ease.

The instructors who have used these materials have found that the six Modules and videos meet many of the needs of the learner. Further, they find that the Modules foster an educational environment that ignites curiosity about the biological world, satisfaction from the mastery of new content and skills, and for some students, an interest in pursuing genetics or bioinformatics research at the undergraduate or graduate level. We have used these Modules in introductory biology courses after the first student exposure to gene structure, in upper-level classes to cement earlier knowledge, and in independent study courses, where students start with the Modules, then pursue independent bioinformatics research projects. The six Modules introduce basic bioinformatics skills in the context of learning about eukaryotic gene structure. Students learn how to use relevant databases and software packages, and gain a deeper understanding of transcription, translation, regulation of gene expression, and genome organization.

We have found that integrating bioinformatics skills acquisition with traditional genomics concepts has broader benefits. Biological research is increasingly dependent on large data sets that require specialized skills to curate and analyze (17,18). There is a growing need for scientists who are comfortable with computer-based activities and have the skill sets needed to effectively use software to process and analyze large data sets.

Intended Audience

The Modules have been used by GEP faculty in Biology courses that include genetics as part of the curriculum. They are intended for and have been used to teach eukaryotic gene structure in a range of different courses (e.g., introductory biology, genetics, cell biology, genomics) at both community colleges and 4-year institutions. The six Modules have been used to teach gene structure and function to undergraduates at all levels (from freshman to senior), and to upper-level high school students taking courses in a community-college setting. Although many GEP faculty teach traditional courses in genetics and genomics, a review of the principles of gene structure and central dogma are often included in courses such as General Biology, Cell Biology, and Developmental Biology, among others; we expect that the Modules could be equally useful in those courses.

The Modules can also be used as stand-alone instructions on how eukaryotic genes are structured. Some faculty have used these materials as beginning instructions in a research-based course to facilitate student participation in the annotation of *Drosophila* genomes as part of the Genomics Education Partnership (<http://gеп.wustl.edu>) research project. The Modules also provide a solid background (or refresher) for students in Bioinformatics courses and Independent Study courses that include gene annotation projects.

Our original goal was to produce a set of self-guided lessons for beginning students (freshmen and sophomores) majoring in the natural sciences who have very little prior knowledge about genes and genomics. The Modules have been used successfully with beginning students, but GEP faculty have found that the Modules were also helpful for students who had already taken a course where gene structure, transcription and translation had been taught. The Modules reinforced biological concepts that these students had already learned, and helped to dispel misconceptions about cellular processes. In addition, exposure to and eventual mastery of additional genomics concepts and terms, such as open reading frame (ORF), start codon, and untranslated regions (UTRs), provided students with the necessary background to understand more complex topics in genetics and genomics.

Required Learning Time

Depending on the level of previous student knowledge, we have found that each Module generally requires 1-2 hours to complete, or a total of 6-12 hours for the entire set. Beginning students sometimes need more than two hours to complete a Module, and instructors usually allow the remaining work to be completed as homework and turned in later.

Most instructors used the Modules during the lab period, which provides a long uninterrupted period (two to three hours) during which students can work. The Modules replaced wet lab experiments. An added benefit of using the Modules in lab is that the number of students in a lab section is usually capped at 18-24, a number that can be effectively managed by an instructor and one or two teaching assistants (TAs).

Each Module can be completed either individually or collaboratively, with students working in pairs or in small groups. We recommend that each student have access to a computer, by bringing their own laptop or having a laptop/desktop provided to them in the classroom or lab. A reliable Internet connection is essential.

Most faculty have used the Modules in the classroom or during a lab session. However, the Modules have also been assigned as homework or independent study. Our experience suggests that students show the most learning gains when they have some guidance from the instructor or a TA as they work on the individual lessons. However, the provided videos cover the topics that we found students needed most help with. Nonetheless, assigning all of the Modules as a homework packet without guidance from the instructor reduces the learning gains, even for advanced undergraduates who have already taken a genetics course.

Some faculty who have taught the curriculum begin classroom instruction with all of the students working on the first task and the first question in a Module. They then proceed to the next task/question only after all students understand and have completed the first task. Other faculty have allowed students to work at their own pace. Whenever a question is problematic for multiple students, a “mini-lecture” can be used to explain the concept to the class. The instructor, and when possible a TA, can serve as a resource by circulating around the classroom to help individual students as needed. For beginning students, the Modules should be taught after the concept of a gene and the basics of transcription and translation have been introduced in lecture.

Preparation time for the instructor using the Modules for the first time is approximately one hour per Module to work through all of the questions and identify concepts that may be problematic for the class. Extra preparation time might be

needed if the instructor wishes to incorporate the videos that accompany the Modules into the course (a strategy that we recommend) or to create “mini-lectures” that explain biological concepts that might be challenging for their students.

Pre-requisite Student Knowledge

To date, the Modules have been used primarily in genetics courses with sophomores or juniors who have completed one or two semesters of general biology. These students are familiar with DNA, RNA, and the concept of a gene, and have a rudimentary understanding of transcription and translation.

The Modules have also been used in a General Biology course with freshmen, who were introduced to genes and gene structure in lecture before starting the Modules as a six-week lab activity. For this group, students were required to read each Module before coming to lab, and were introduced to the main topics by the instructor before they started working.

Pre-requisite Teacher Knowledge

The instructor should have a good understanding of the structure of a gene, how conserved sequence motifs (e.g., core promoter motifs, transcription factor binding sites) facilitate protein-nucleic acid interactions, and the mechanisms of transcription, translation, and alternative splicing. The instructor does not need to have a comprehensive background in bioinformatics. The Modules integrate conventional biology topics with which the instructor will already be familiar, with a computer-based approach for discovering and modeling the structure of a gene from *Drosophila*. In addition, instructors should understand how to use the basic navigation features of the UCSC Genome Browser; this understanding can be acquired by working through the first Module, including watching the associated video. Instructors should work through each Module before assigning the Module to students, especially if they have limited prior experience using the UCSC Genome Browser.

SCIENTIFIC TEACHING THEMES

Active learning

Because the Modules can be used in many different contexts, the instructor could use multiple active-learning approaches. However, in most of our implementations, students work in groups of two or three using their own computers or computers in a computer lab. The instructor gives a brief introduction to the Module, and then students discuss each part of the lesson within their small groups while they explore using the browser. We also often use large group discussions of student questions and challenging concepts. Typically, students work together to complete each Module and submit their answers to the questions associated with each Module to their instructor. This is consistent with the classic definition of active learning as “anything that involves students in doing things and thinking about the things they are doing” (19). Active learning often includes higher order thinking tasks such as analysis, evaluation and synthesis, and this may be especially true for group work (6, 19). As students discuss the exercises within each Module, they gather evidence, evaluate potential answers to each question, and occasionally resolve contradictions in the evidence.

Using a group format has some risks of uneven participation, but generally stimulates useful dialogue. Some instructors have used the Modules as an outside-of-class independent project, in which students work through each Module independently,

watching the videos that accompany each Module, and answering the questions in each Module for submission.

Assessment

Each Module contains multiple embedded questions to assess student knowledge gains. If students can answer the questions correctly, they have completed the task assigned in each Module. Some faculty give credit for completion of the Modules without evaluating individual questions, while others grade the embedded questions in each Module. Students can also be given instructor-designed pre- and post- quizzes to assess knowledge gains from the use of the Modules. An answer sheet for each Module is provided in the Supporting information (Supporting files S13-S18). Pre- and post-quizzes and Module answer keys are available upon request.

Inclusive teaching

These Modules make use of multiple approaches to facilitate learning about gene structure and using the genome browser. Students can watch videos that explain concepts such as splicing and phase, and explain use of a bioinformatics tool (e.g., the genome browser). Each Module also has short readings to explain key terms and concepts, and requires students to use the genome browser to visually explore DNA sequence and other genomic features to study gene structure. While there are a number of approaches that instructors have used in the classroom, the most often used is a team-based, cooperative approach where students work together to understand eukaryotic gene structure and produce a gene model by the end of the Module Six. In these classrooms, instructors work closely with student teams to ensure that all groups of students are included and that all students meet the learning objectives.

The Modules also expose students to an important bioinformatics research tool, the genome browser, and uses it as a pedagogical tool. The UCSC Genome Browser, the GEP UCSC Genome Browser mirror, and others, are publicly available online. The Modules provide training in the use of these freely available tools to understand complex biological phenomena. They also train students so that they are ready to participate in a course-based research experience involving gene annotation. The Modules have been successfully implemented in community colleges and 4-year colleges/universities, where students have a wide range of levels of preparation.

LESSON PLAN

Prior to using the Modules, students should have received an introduction to the Central Dogma, DNA structure, RNA structure, protein structure, transcription, and translation. The Modules will allow students to study gene structure, mRNA processing, and the production of gene products by studying the different datasets that can be visualized through the Genome Browser, using the browser to “zoom in” for details as needed.

Specific lesson plans are provided on the first page of each Module in bullet form and are also described below. These Modules can be used for independent study, but we have found that using them in the classroom with the instructor and other students present is the best approach in an introductory context. Each Module is available as a Microsoft Word document or PDF file at the “Introducing Genes” page of the GEP web site (http://gеп.wustl.edu/curriculum/introducing_

genes) along with associated videos and a glossary of terms, under the heading “Understanding Eukaryotic Genes.” We typically provide students with printed copies of each Module at the beginning of the period in which they are used.

The Modules focus on the structure of the *tra* gene of *Drosophila melanogaster*, an alternatively spliced gene required for development and sex determination. Some Modules also focus on the structure of nearby genes (i.e., *CG32165* and *spd-2*), which are also visible in the genome browser view provided.

Module 1: What is a gene

We start with “Module 1: Introduction to the Genome Browser: What is a Gene?” Instructors may wish to start with a class discussion on “what is a gene?” including discussions of the functions of a gene and how these functions are related to gene structure. Students should then work through the computer exercise that introduces the genome browser and the genomic features that can be highlighted using the genome browser. Students should also watch the associated videos on the genome browser and on evidence tracks. The class should stop at different points (selected by the instructor) to discuss what they have learned and to address questions from students.

Module 2: Transcription Part I

Module 2 describes transcription. In Module 2, we have found it best to start with a class discussion of the process of transcription. The discussion topics include: What is transcription? What cellular proteins are required for transcription? How does it work mechanistically? What is/are the products of transcription? Students then work through the Module. The Module describes signals that regulate transcription and illustrates how to use the Short Match functionality of the genome browser to search for the transcription start site (TSS) and the transcription termination sequence. (A video associated with this Module illustrates the Short Match functionality.) Students identify where transcription starts and ends for the *tra* gene in *Drosophila melanogaster*, and identify the length of the primary transcript. They also examine RNA-Seq data and use it to study the transcript, and can watch a video on RNA-Seq to support their learning. Instruction using this Module concludes with a class discussion or a homework assignment that addresses the following questions: How important is it for RNA polymerase II to recognize the promoter sequence? Do you think it is possible for a gene to have more than one TSS? How would RNA polymerase II “know” which TSS to choose? When would different transcription start sites make a difference in the protein product, and when not?

Module 3: Transcription Part II

We have found it most effective for students to begin Module 3 by discussing the following questions: What happens to the initial (pre-mRNA) transcript made by RNA pol II? Does it leave the nucleus “as is” or do changes to the pre-mRNA have to occur first? The instructor then gives a mini-presentation illustrating the 5’ capping, 3’ polyadenylation, and splicing of introns that occurs during pre-mRNA processing. Students then begin to work through the Module. The instructor can either pause to discuss the answers to the questions or ask the

students to turn in the answers for evaluation. The instructor can conclude this Module with a wrap-up discussion on mRNA processing.

Module 4: Splicing

The instructor could begin Module 4 with a review of mRNA processing, introduce splicing and the spliceosome, and introduce the term isoform. The instructor should ask students to watch the “Genes and Isoforms” video that accompanies this Module. Students should also familiarize themselves with RNA-Seq data by watching the associated “RNA-Seq and TopHat” video, and complete Investigation 1 of the Module. The instructor will then introduce students to the concept of a consensus sequence so that they can complete Investigation 2 of the Module by locating the splice donor and acceptor sites for the first intron of *tra*-RA. Using the information provided, students will locate the splice donor and acceptor sites for the second intron in Investigation 3 of the Module. Students then discuss the length of the pre-mRNA compared to the length of the spliced mRNA, and finally identify isoforms with different transcription start sites or alternative splicing patterns.

Module 5: Translation

The instructor could begin Module 5 with a review of translation, giving students an overview of the ribosome, tRNAs, and associated proteins involved in translation, as well as a review of the genetic code. Students will then work through the activities of the Module using the genome browser. The class can either pause to discuss answers to embedded questions, or the answers can be turned in for assessment. As students work through the splicing section of the Module, they will learn about the concept of the phase of the splice donor and acceptor sites. Many students initially might find the concept of phase to be confusing, specifically the fact that an exon can end in the middle of a codon. The instructor can ask the students to watch the associated “Splicing and Phase” video followed by a class discussion to clarify this concept. A concluding discussion will emphasize the following: (1) translation of mRNA into amino acids using triplet codons; (2) identification of open reading frames; (3) maintenance of open reading frames across splice sites to generate an mRNA that produces a working protein product; (4) the assembled open reading frame must begin with a start codon and end with a stop codon. The use of the browser makes identification of open reading frames very quick and visual (stop codons are marked in red), and quickly establishes the fact that the reading frame can shift from one exon to the next.

Module 6: Alternative Splicing

In Module 6, students learn about the concept of alternative splicing by considering the two alternatively-spliced isoforms of *tra*: *tra*-RA and *tra*-RB. The instructor should give a mini-lecture that introduces students to the role of *tra* in sex determination and the differences between the two isoforms of *tra*. A sample lecture is provided (Supporting file S20). Students will encounter many concepts (e.g., reading frames, phase) that were introduced in prior Modules. It might be helpful for students to re-watch the “Genes and Isoforms” video, the “Splicing and Phase” video, and the “RNA-Seq and

TopHat” video to review these important concepts. Students should then work through Investigation 1 to compare the RNA-Seq expression patterns in the adult males and adult females samples and to understand how different mRNAs can be encoded in the same gene. Students then work through Investigation 2 to examine how alternative splicing could result in the production of different polypeptides. As part of this Investigation, students will construct a gene model for *tra*-RB using sequence information and RNA-Seq data as evidence. The Module concludes with a discussion of the impact of alternative splicing on protein function and how the structures of the two isoforms of *tra* correspond to the difference in RNA-Seq read coverage in the adult males and adult females samples. Alternative splicing of *tra* determines the secondary sexual characteristics of the fly, a fact that can impress on students the importance of correct splicing.

Lesson Plan Timeline

Please see Table 1.

TEACHING DISCUSSION

Implementation

Nine faculty from nine different institutions have used these Modules to teach gene structure during Fall 2015 to Spring 2016. Seven of the faculty used all six Modules in the order presented. Two faculty have used only one or a few of the Modules, as we have also found that “one size does not fit all” is true even for beginning students (20). Module 1 has been used as an introduction to or a review of the UCSC Genome Browser. Modules 1 and 3-5 have been used as a review of transcription and translation, in the context of a eukaryotic gene. Other configurations may be used, and these variations reflect the needs of the instructor, students, and course content. Lesson plans may be customized to reflect a faculty member’s course objectives, or expertise with a different model organism. For example, using the Modules as a template, two GEP faculty have written brief lesson plans to investigate the human *parkin RBR E3 ubiquitin protein ligase* gene, and illustrated biological concepts using examples of genes from a model organism used in their own research. This does require some additional preparation time on the part of the instructor. However, the instructor can use the different Modules as a template and then incorporate genes or content from their course.

Specialized training workshops are not needed for faculty to teach the materials in these Modules. However, faculty interested in bioinformatics education and research can attend any of a number of workshops to gain these skills. In addition to the resources mentioned in the Introduction, RNA-Seq and genome browser workshops are hosted by university entities such as the University of California, Davis Bioinformatics Core (<http://bioinformatics.ucdavis.edu/training/>), and professional development courses are available through Bioinformatics.org. Cost does not have to be a barrier; EMBL-EBI offers free course materials online (<https://www.ebi.ac.uk/training/online/>), and past Canadian Bioinformatics Workshops are available under a Creative Commons License (<https://bioinformatics.ca/workshops>).

Effectiveness in Meeting Learning Objectives and Student Reactions

Although formal assessment has not been completed, instructors report learning gains using pre- and post- quizzes covering the six Modules. Formal assessment of the curriculum is underway, and anecdotal reports from participating faculty indicate that students show gains in understanding transcription and translation, and in using the genome browser. In addition, students in the classes in which the curriculum was implemented generally have favorable impressions of their learning. Through informal discussions, we have determined that the curriculum noticeably reduces student frustration that may be associated with integrating content and skills. Students show confidence in their ability to answer questions in the Modules. In addition, students are very comfortable using their own laptops, displaying a sense of control over their learning environment (choosing which tools to display in various tabs and how quickly to move through the exercises, for example).

A Bridge to Independent Research

These Modules use a multidisciplinary approach to help students learn about the structure of eukaryotic genes while developing skills in using the genome browser. The Modules use real biological data from *D. melanogaster* and publicly available bioinformatics research tools to build transferable skills that could be used for more advanced projects(16). Some faculty have used the Modules to help prepare students to participate in gene annotation of a *Drosophila* genome, part of a larger collaborative research endeavor by the GEP. As a result, approximately 1000 students each year across the US, Taiwan and Canada annotate recently sequenced *Drosophila* genomes as part of an ongoing effort to understand the evolution of the Muller F element and the factors that regulate gene expression in this heterochromatic environment (21).

SUPPORTING MATERIALS

- S1. Module 1: What is a Gene?
- S2. Module 2: Transcription Part I
- S3. Module 3: Transcription Part II
- S4. Module 4: Splicing
- S5. Module 5: Translation
- S6. Module 6: Alternative Splicing

FILES S7-S12 can be found at <https://drive.google.com/drive/folders/0B33BU08B2owHUHJ0RkNuYl83YOU?usp=sharing>

- S7. Browser video
- S8. Genes and isoforms video
- S9. RNA-Seq and TopHat video
- S10. Short Match video
- S11. Splicing and Phase video
- S12. Tracks video

- S13. Module 1: What is a Gene? Answer Sheet
- S14. Module 2: Transcription Part I Answer Sheet
- S15. Module 3: Transcription Part II Answer Sheet
- S16. Module 4: Splicing Answer Sheet
- S17. Module 5: Translation Answer Sheet
- S18. Module 6: Alternative Splicing Answer Sheet
- S19. Glossary of terms
- S20. Sex determination mini-lecture

ACKNOWLEDGMENTS

Supported by NSF IUSE grant #1431407 to SCRE.

REFERENCES

1. Bloom BS. 1984. The 2 Sigma problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educ Res* 11:3-4-16.
2. Lopatto D. 2007. Undergraduate Research Experiences Support Science Career Decisions and Active Learning. *CBE Life Sci Educ* 6:297-306.
3. Lopatto D. 2009. Science in Solution: The Impact of Undergraduate Research on Student Learning. Research Corporation for Science Advancement.
4. Gasiewski JA, Eagan MK, Garcia GA, Hurtado S, Chang MJ. 2012. From Gatekeeping to Engagement : A Multicontextual , Mixed Method Study of Student Academic Engagement in Introductory STEM Courses. *Res High Educ* 53:229-261.
5. Froyd JE. 2005. White Paper on Promising Practices in Undergraduate STEM Education.
6. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Pat M. 2014. Active learning increases student performance in science , engineering , and mathematics. *Proc Natl Acad Sci* 111:8410-8415.
7. Haak DC, HilleRisLambers J, Pitre E, Freeman S. 2011. Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science* (80-) 332:1213-1216.
8. Michael J. 2006. Where 's the evidence that active learning works ? *Adv Physiol Educ* 30:159-167.
9. Roberts JA, Honore LUC, Yaya P, Manolis C. 2014. The Invisible Addiction : Cell-phone Activities and Addiction Among Male and Female College Students. *J Behav Addict* 3:254-265.
10. Revere L, Kovach J V. 2011. Online Technologies For Engaged Learning: A Meaningful Synthesis for Educators. *Q Rev Distance Educ* 12:113-124.
11. Buttigieg PL. 2010. Perspectives on Presentation and Pedagogy in Aid of Bioinformatics Education. *Brief Bioinform* II:587-597.
12. Smith MK, Wood WB, Adams WK, Wieman, C, Knight JK, Guild N, Su TT. 2009. Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science* (80-) 323:122-124.
13. Awong-Taylor J, D'Costa A, Giles G, Leader T, Pursell D, Runck C, Mundie T. 2014. Undergraduate Research for All: Addressing the Elephant in the Room. *CUR Q* 37.
14. Wood L, Gebhardt P. 2013. Bioinformatics Goes to School -- New Avenues for Teaching Contemporary Biology. *PLoS Comput* 9:e1003089.
15. Shaffer CD, Alvarez C, Bailey C, Barnard D, Balla S, Chandrasekaran C, Chandrasekaran V, Chung H, Dorer DR, Du C, Eckdahl TT, Poet JL, Frohlich D, Goodman AL, Gosser Y, Hauser C, Hoopes LLM, Johnson D, Jones CJ, Kaehler M, Kokan N, Kopp OR, Kuleck GA, Mcneil G, Moss R, Myka JL, Nagengast A, Morris R, Overvoorde PJ, Shoop E, Parrish S, Reed K, Regisford EG, Revie D, Rosenwald AG, Saville K, Schroeder S, Shaw M, Skuse G, Smith C, Smith M, Spana EP, Spratt M, Stamm J, Thompson JS, Wawersik M, Wilson BA, Youngblom J, Leung W, Buhler J, Mardis ER, Lopatto D, Elgin SCR. 2010. The Genomics Education Partnership : Successful Integration of Research into Laboratory Classes at a Diverse Group of Undergraduate Institutions. *CBE-Life Sci Educ* 9:55-69.
16. Ditty JL, Kvaal CA, Goodner B, Freyermuth SK, Bailey C, Robert A, Gordon SG, Heinhorst S, Reed K, Xu Z, Sanders-lorenz ER, Axen S, Kim E, Johns M, Scott K, Kerfeld CA. 2010. Incorporating Genomics and Bioinformatics across the Life Sciences Curriculum. *PLoS Biol* 8:e1000448.
17. Rosenwald AG, Pauley MA, Welch L, Elgin SCR, Wright R, Blum J. 2016. The CourseSource Bioinformatics Learning Framework. *CBE Life Sci Educ* 15:1-2.
18. Cummings MP, Temple GG. 2010. Broader incorporation of bioinformatics in education : opportunities and challenges. *Brief Bioinform* II:537-543.
19. Bonwell CC, Eison JA. 1991. Active Learning: Creating Excitement in the Classroom.
20. Ranganathan S. 2005. Bioinformatics Education--Perspectives and Challenges. *PLoS Comput Biol* 1:e52.
21. Leung W, Participating Students and Faculty of the Genomics Education Partnership. 2015. *Drosophila Muller F Elements Maintain a Distinct Set of Genomic Properties Over 40 Million Years of Evolution*. *G3 Genes Genomes Genet* 5:719-740.

Table 1. Lesson Timeline

Activity	Description	Time	Notes
Module 1. Introduction to the Genome Browser: What is a Gene?			
Warm up	Discuss the question: What is a gene? (Perhaps discuss with a partner, then as a class, or start with whole class discussion.) Emphasize the function of a gene; consider how the structure of the gene is related to its function.	10 minutes	
Investigation	1. Watch Genome Browser video at https://youtu.be/suMC7wmP7tA . 2. Work through Module 1, stopping on page 4 to watch the Evidence Tracks video at https://youtu.be/BwwP7cOqr0Y .	50 minutes-1 hour 40 minutes	Module 1 is provided as Supporting File S1
Conclusion	Conclude with discussion of the following points: <ul style="list-style-type: none"> • Genes may run in either direction on a chromosome; • Genes are represented on the genome browser as blocks connected by lines; • Eukaryotic genes are made up of protein-coding exons (the blocks) connected by introns; • Proteins usually begin with a Methionine (M) and end at a stop codon (*) 	15 minutes	
Module 2. Transcription Part I			
Warm up	1. Discuss the questions: What is transcription? What cellular proteins are required for transcription? How does it work mechanistically? What is/are the products of transcription? (students discuss in pairs, then as a class) 2. Describe regulatory signals: Transcription Start Site (TSS), and AATAAA sequence (site of transcript cleavage for termination).	10 minutes	
Investigation	Work through Module 2, stopping at page 4 to watch the RNA-Seq and TopHat video at https://youtu.be/qepVXEsfLMM and at page 10 to learn about the Short Match functionality of the Genome Browser at https://youtu.be/eoeWufgcdvg	50 minutes - 1.5 hours	Module 2 is provided as Supporting File S2
Conclusion	Conclude by challenging students to think about these questions: <ol style="list-style-type: none"> 1. How important is it for RNA polymerase II to recognize the promoter sequence? 2. Do you think it is possible for a gene to have more than one transcriptional start site? 3. How would RNA polymerase II know which one to choose? 4. When would it make a difference in the protein product, and when not? 	15 minutes	
Module 3. Transcription Part II			
Warm up	1. Discuss the questions: What happens to the initial (pre-mRNA) transcript made by RNA pol II? Does it leave the nucleus 'as is'? Or do changes have to occur? (Hint: introns vs. exons) 2. Mini-presentation illustrating that during pre-mRNA processing, three events occur: <ul style="list-style-type: none"> - 5' capping, - 3' polyadenylation, - splicing out of introns 	10 minutes	
Investigation	Work through Module 3.	50 minutes - 1.5 hours	Module 3 is provided as Supporting File S3
Conclusion	Conclude with a discussion of mRNA processing, emphasizing the three steps: <ul style="list-style-type: none"> • 5' capping, • 3' polyadenylation, • removal of introns through splicing (via spliceosome) 	15 minutes	

Activity	Description	Time	Notes
Module 4. Splicing			
Warm up	1. Review mRNA processing. 2. Introduce splicing and the spliceosome.	10 minutes	
Investigation	1. Work through Module 4, Investigation 1. Watch RNA-Seq and TopHat video at https://youtu.be/qepVXEsfLMM . When students work on page 6 of the Module, they should also watch the Genes and Isoforms video at https://youtu.be/8jtTp_6vN4M . 2. Work through Investigations 2 and 3.	50 minutes-1.5 hours	Module 4 is provided as Supporting File S4
Conclusion	1. Discuss the length of the pre-mRNA vs. the length of the spliced mRNA 2. Identify isoforms of <i>tra</i> with different transcription start sites or alternative splicing patterns.	15 minutes	
Module 5. Translations			
Warm up	1. Review the process of translation: Overview of the ribosome, tRNAs, and associated proteins involved in translation (Initiation Factors, Elongation Factors and Release Factors). 2. Review genetic code table.	10 minutes	
Investigation	1. Work through Module 5 2. When working through page 8 of the Module, watch the Splicing and Phase video at https://youtu.be/JsvUfHy3eHE .	50 minutes-1.5 hours	Module 5 is provided as Supporting File S5
Conclusion	Discuss the following points: <ul style="list-style-type: none"> • mRNAs are translated into amino acids using triplet codons • Identification of ORFs • The ORF must be maintained across splice sites to generate a working mRNA • The assembled ORF begins with a start codon and ends with a stop codon. 	15 minutes	
Module 6. Alternative Splicing			
Warm up	Introduce students to <i>tra</i> -RB, the second isoform of <i>tra</i> .	10 minutes	
Investigation	1. Work through Module 6 2. When working through page 2 of the Module, students may wish to rewatch the RNA-Seq and TopHat video at https://youtu.be/qepVXEsfLMM . 3. When working through page 4 of the Module, students may wish to rewatch the Splicing and Phase video at https://youtu.be/JsvUfHy3eHE . 4. When working through page 6 of the Module, students may wish to rewatch the Genes and Isoforms video at https://youtu.be/8jtTp_6vN4M .	50 minutes-1.5 hours	Module 6 is provided as Supporting File S6
Conclusion	1. Discuss gene models. 2. Discuss the following points: How does the polypeptide translated from the <i>tra</i> -RB isoform differ from the polypeptide translated from the <i>tra</i> -RA isoform? What are the consequences of these differences on protein function? 3. Discuss how the bigger mRNA leads to creation of a smaller polypeptide! 4. Consider how alternative splicing could allow many different proteins to be encoded by the same gene. 5. Based on the gene structure of the two isoforms of <i>tra</i> shown in the “FlyBase Genes” track, provide a hypothesis that could explain this difference in RNA-Seq read coverage between the adult males sample and adult females sample.	15 minutes	