How to Find a Gene: Retrieving Information From Gene Databases

Ah Rume Julie Park[1,2†], Maitreyi Upadhyay[1,2†], William J. Anderson[1,2], and Amie L. Holmes[1,2*]

[1] Department of Stem Cell and Regenerative Biology, Harvard University

[2] Harvard Stem Cell Institute

[†] These authors contributed equally to the work.

Abstract

A strong understanding of distinct gene components and the ability to retrieve relevant information from gene databases are necessary to answer a diverse set of biological questions. However, often there is a considerable gap between students' theoretical understanding of gene structure and applying that knowledge to design laboratory experiments. In order to bridge that gap, our lesson focuses on how to take advantage of readily available gene databases, after providing students with a strong foundation in the central dogma and gene structure. Our instructor-led group activity aids students in navigating the gene databases on their own, which enables them to design experiments and predict their outcomes. While our class focuses on cardiomyocyte differentiation, classes with a different focus can easily adapt our lesson, which can be conducted within a single class period. Our lesson elicits high engagement and learning outcomes from students, who gain a deeper understanding of the central dogma and apply that knowledge to studying gene functions.

Citation: Park ARJ, Upadhyay M, Anderson WJ, Holmes AL. 2023. How to Find a Gene: Retrieving Information From Gene Databases. CourseSource 10. https://doi.org/10.24918/cs.2023.8

Editor: Rou-Jia Sung, Carleton College

Received: 9/3/2021; **Accepted:** 12/28/2022; **Published:** 2/23/2023

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

*__Correspondence to:__ Harvard University, 7 Divinity Avenue, Fairchild G55, Cambridge Massachusetts 02138, USA, amie_holmes@harvard.edu

Learning Goals

Students will:

◊ learn how to retrieve gene information from genomic databases.

◊ apply their knowledge of gene structure, transcripts, and protein products to design a gene expression experiment.

◊ From Biochemistry and Molecular Biology Learning Framework:

  » What is a genome?

◊ From Bioinformatics Learning Framework:

  » Where are data about the genome found (*e.g.,* nucleotide sequence, epigenomics) and how are they stored and accessed?

  » How can bioinformatics tools be employed to analyze genetic information?

  » Where are data about the transcriptome found (*e.g.,* expression, epigenomics and structure) and how are they stored and accessed?

◊ From Genetics Learning Framework:

  » How is DNA organized?

  » What are the molecular components and mechanisms necessary to preserve and duplicate an organism's genome?

  » What experimental methods are commonly used to analyze gene structure and gene expression?

Learning Objectives

Students will be able to:

◊ navigate gene databases and

  » locate gene structure information (*e.g.,* exon number, gene size, chromosome location, etc.).

  » extract transcript expression patterns.

  » predict protein sizes.

◊ use the collected gene information to design and execute experiments, such as PCR and qRT-PCR.

◊ predict changes in gene expression during cellular differentiation (in this exercise, from pluripotent stem cells to cardiomyocytes).

INTRODUCTION

Our understanding of "the anatomy of a gene" has evolved with the advent of new technologies, which allows us to probe genes and their structures in more depth and detail than ever before. Correspondingly, the use of gene databases enables systematic storage, retrieval, and processing of that information. Our views on gene structure therefore need to transition from "static" to "dynamic," allowing for broader applications of such information.

Introductory information on gene features, including structure, directionality, and regulatory regions, is largely generalized. However, application of such knowledge often occurs in specific contexts, such as designing primers for sequencing or guide RNAs for CRISPR-mediated gene targeting, and predicting band sizes in a Western blot experiment to confirm a gene knockout. There is often a gap when students attempt to apply their general knowledge about gene structure to design experiments in a laboratory setting, which requires retrieval of specific and context-dependent information. Our lesson provides a holistic experience for students to better understand gene architecture by learning how to retrieve gene structure information from gene databases and use that knowledge to design and predict experimental outcomes.

Genomic and bioinformatics analyses are recognized as integral parts of undergraduate science education, and thus, there are a growing number of courses that focus on introducing interactive lessons on genome browser navigation (1–5). In such courses, web-based tools, such as the NCBI, Ensembl, and UCSC genome browsers, are often introduced in the context of learning about the central dogma, nucleotide sequences, introns, and exons (6). The same tools are also used to construct primers and extract gene information for designing experiments and predicting experimental outcomes (7). Furthermore, web-based platforms allow instructors to generate course-specific genome browsers for Course-based Undergraduate Research Experiences (CUREs) that focus on efficient data visualization and genome annotation (8). Bioinformatics CUREs like the Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) integrate research efforts with curriculum, thereby uncovering new insights about the diversity of the bacteriophage population while introducing undergraduate students to discovery-based research experiences (9). As a reflection of the growing interest in the incorporation of data analysis in undergraduate curricula, a collective effort to centralize interactive teaching material for data analysis has been introduced (10). This hub provides a continuously evolving collection of tutorials for undergraduate and graduate students. Altogether, these interactive exercises demonstrate the importance of providing students with a hands-on experience to learn and solidify their understanding of genes and gene structures.

Within *CourseSource* specifically, Laakso *et al.* (11) provides a comprehensive, six module bioinformatics curricula using the UCSC genome browser that focuses on gene expression and its usage in accurate genomic annotation. Comparable to Laakso *et al.*, our single-class period lesson provides real-time, hands-on guidance on using gene databases to retrieve gene structure information and the application of this knowledge to gain a deeper understanding of gene regulation and experimental design. Toward that goal, each module in Laakso *et al.* and our lesson have a similar format and are designed to be completed within a 1–2 hour class period. Our class provides an extended instructor-led session where the students first explore the genome browser in the context of a specific human gene, *SOX2*, along with the instructor and then proceed into the investigation session by working on an in-class group activity to explore another gene of interest. For the investigative session, students are given time to work on handout questions together and discuss key concepts sitting in small groups (12). The instructional teams checked on student progress, answered questions throughout the session, and probed the students further with additional questions (13). During our lesson, we also encouraged students to simultaneously use the whiteboard to draw out different genomic architecture observed while navigating through the genome browser. These student drawings served as visual cues while reviewing answers and central concepts in the concluding session. While Laakso *et al.* use RNA-seq data analysis as a means to understand gene expression, our lesson focuses on designing experiments that involve techniques such as reverse transcription-quantitative polymerase chain reaction (RT-qPCR) (designing primers) and Western blot (predicting protein size) to apply what they have learned in different genomic contexts. Students had the opportunity to engage more deeply with the material using additional questions. These questions served as another learning activity in itself, where students engaged in higher-order thinking with more time (14).

In sum, our single-class period exercise provides a complementary and alternative lesson to previously published exercises, where students obtain guided exposure to using the NCBI and Ensembl databases, and then apply what they have learned in different genomic contexts and experimental design.

Intended Audience

This lesson was used in a sophomore-level introductory molecular biology course (that included some juniors and seniors) at a large private research university. Most students were pursuing a life science major and some students had experience conducting independent laboratory research. This exercise can easily be used by community colleges and primarily undergraduate institutions given that the genome databases are publicly available and no special equipment or reagents are necessary for implementation.

Required Learning Time

The lesson takes 75 minutes. Students may spend extra time working on the optional after-class handout and attending office hours to ask questions.

Prerequisite Student Knowledge

Before engaging in this lesson, students should have completed introductory biology and chemistry courses that cover the foundational principles of molecular biology, cell biology, genetics, genomics, evolution, and chemistry. Especially in introductory biology, students should have been exposed to the concepts of gene structure (exons, introns, promoters, enhancers), transcription (transcription start and stop sites, untranslated regions, alternative splicing), and translation. It would be helpful for students to know the basics of polymerase chain reaction and blotting.

Prerequisite Teacher Knowledge

The instructor should be familiar with gene structure and the central dogma. The instructor should familiarize themselves with the different features of, and how to navigate, the NCBI Gene and Ensembl databases. NCBI has a series of how-to tutorial videos on YouTube that students and instructors may find helpful: A Beginner's Guide and Overview.

SCIENTIFIC TEACHING THEMES

Active Learning

During the instructor-led genome browser navigation, the instructor asks questions and elicits answers from students to maintain a high level of engagement. Students sit in pairs within groups of four, and each group illustrates their answers and ideas on a whiteboard. This setting allows students to engage in spontaneous think-pair-share and group discussions throughout the session.

Students work on the in-class group activity and optional after-class handout, which requires students to actively revisit and apply their knowledge from navigating gene databases.

Assessment

In-class group activity (formative)

The in-class group activity (Supporting File S1, slide 30) allows students to self-evaluate their learning by checking their progress with other classmates while completing tasks using the genome browser (see Supporting Files S1 [slides 31–34] or S2 for answers). During the activity, instructors walk around the room answering and asking questions, while gauging student understanding and identifying steps/concepts which students are having difficulty grasping.

Optional after-class handout (formative)

This resource extends the in-class exercise by asking students to explore different genes and answer additional questions using the genome browser (Supporting Files S3, S4). This allows students to review what they have learned in class, and to ascertain their understanding of genome browsers to extract information. This handout is not graded and enables students to either self-evaluate their understanding by completing the handout on their own and checking their answers using the answer key, or by attending office hours to ask questions about the handout. For students who come to office hours, instructors can identify concepts that students are having trouble understanding and help them navigate the genome browser one-on-one.

This exercise provides a context for what the students will do in the two upcoming laboratory sections, namely performing a RT-qPCR and a Western blot to measure the expression of various genes at three different time points in the differentiation of pluripotent stem cells to cardiomyocytes. Students are asked to use the genome browser to gather more information about these genes. By looking up their function, students can formulate hypotheses on how gene expression levels change during development. Furthermore, gene structural information can be used for designing primers for RT-qPCR and for predicting band sizes on Western blots. These exercises allow students to "think like a scientist" as they design and predict outcomes of experiments.

Take-home exam (summative)

Students apply what they learned in class to a new context in the take-home exam (Supporting Files S5, S6). Students are encouraged to collaborate with other classmates but to compose their own answers. Students have access to the genome browsers and lecture material, and can utilize any other resources with proper citation. In this assignment, students convey their understanding of how to use and interpret information from the genome browser by extracting structural information about a gene, designing an experiment using information from the genome browser, and comparing wild-type and mutant gene structure highlighting the consequence of the mutation. This assessment is assigned after the in-class exercise and graded.

Inclusive Teaching

The lesson is designed to engage all students with varying levels of preparation by beginning with a refresher on concepts from the central dogma to introduce the material. During the refresher and throughout the lesson, we incorporate multiple modes of learning (*e.g.,* visual, auditory, hands-on) and classroom settings to facilitate student learning.

The instructor points to specific parts of the gene illustrations on the slides during the refresher to provide both visual and verbal descriptions. Furthermore, during the instructor-led genome browser navigation, specific parts from the screenshots of the genome browser are highlighted. This allows students to follow the instructor in real time and confirm their work as they navigate the genome browser themselves. Slides and an audio recording of the session are uploaded after class for students who wished to revisit the material at a later time.

For adapting the lesson for visually-impaired students, extra resources should be provided for retrieving gene information from the databases, including extensive written descriptions or verbal descriptions of each step of the procedure from an instructional team member. For students with hearing impairment, screenshots of the genome browser should be supplemented with labels that point to specific parts and extensive written descriptions of the procedures. Once gene information is retrieved, however, students can use the information to move forward with designing experiments on their own.

For students with limited access to computers, the lesson could be done with students working individually with other web-enabled devices (*i.e.,* tablets or smart phones) along with instructor-provided screenshots of different steps in the exercise. Should there be a limited number of web-based devices, students can work in groups.

We paid particular attention to classroom settings in order to allow efficient group work and check-ins by the instructional team. When navigating the genome browser, students can check their progress with their peers. While working on the in-class group activity on the *NANOG* gene, students are encouraged to collaborate and write their answers on a whiteboard in their small groups, in an effort to help reduce anxiety prior to sharing answers with the entire class and to potentially increase overall student learning. For classrooms without whiteboards, students can use paper for illustrating and sharing their ideas during the group exercise.

This lesson also provides an opportunity to discuss the sources and ethics of human genomics data. While our lesson did not address this topic, we suggest some discussion points that may be integrated into the lesson in the Modifications section.

LESSON PLAN

This activity is designed as an in-class application exercise after completing lectures related to the central dogma of biology, including gene structure, transcription, and translation. Furthermore, the examples and questions in this exercise are tied to a RT-qPCR laboratory experiment, but this exercise can easily be separated from the lab and done on its own. During the lab, students utilize RT-qPCR to investigate marker gene expression at different time points in the differentiation of stem cells to cardiomyocytes. Thus, it is helpful for students to have an understanding of PCR and primer design as PCR primer locations will be identified and discussed in the exercise. Since our course focuses on molecular biology through the lens of developmental and stem cell biology, we explore the structure of pluripotency and cardiomyocyte-specific genes. This exercise can be easily adapted to different biological contexts and genes of interest.

The genome browser class is broken up into four parts; (i) a refresher on the central dogma and gene structure, (ii) an instructor-led genome browser navigation, (iii) an in-class group activity, and (iv) presentation of the in-class group activity (Table 1). The class begins with a brief recapitulation of gene structure and information that is necessary for determining gene, transcript, and protein structures by discussing gene directionality, exons, introns, untranslated regions (UTRs), coding sequences (CDSs), and protein sequences (Supporting File S1). Instructors can then pose the question, “How do we use this gene information in a research setting?” providing students with the opportunity to discuss the importance of gene information in designing PCR primers, predicting protein size for western blots, generating gene knockouts, creating expression vectors, etc. In our exercise, we focused on PCR primer design and explained the premise of an RT-qPCR experiment and how we can use this technique to measure gene expression changes in the differentiations from human stem cells to cardiomyocytes. To highlight the basic information we can obtain from gene databases, we explored the structure of a human gene—for example, *SOX2*, which is a pluripotency gene with a simple gene structure—as an entire class. Using the provided human *SOX2* gene (1…2512 nucleotide [nt]), mRNA (1…2512 nt), coding sequence (CDS) (437…1390 nt), and protein length (317 amino acid [aa]) information, we build out a model to reflect the gene, transcript, and protein structure in real-time. To build the *SOX2* gene structure model, we can first deduce that *SOX2* has one exon, because the mRNA transcript is the same length as the gene, both 2512 nt. Since the CDS spans from 437 nt to 1390 nt, we can then build in the 5′ and 3′ UTRs into our model which span from 1–436 nt for the 5′ UTR and 1391–2512 nt for the 3’UTR. Lastly, considering the CDS is 954 nt long, the protein length will be 317 aa which can be calculated by subtracting 3 nt for the stop codon from the 954 nt CDS and then dividing by 3 nt for amino acid codon length. Building this gene model for *SOX2* in real-time allows students to connect the provided gene information with gene structure. The logical question that arises, however, is where the provided gene information for *SOX2* came from, which creates a nice transition into the instructor-led genome browser navigation section.

After building a model of human *SOX2* with the provided information, a step-by-step guide is given on where to find this gene information in two online databases, NCBI Gene database and Ensembl database. As screenshots from the NCBI and Ensembl databases are projected on the PowerPoint slides, the students perform the same tasks of searching the databases in real-time on their computers. For our exercise, we focus on human genes, but we highlight that many different species are available in the gene databases. In NCBI, students are shown where to find summary gene information, gene maps, and detailed gene sequences, how to decipher accession numbers, and how to locate gene position and enhancer information. While there is overlap in the data provided in the NCBI and Ensembl databases, we highlight that Ensembl provides more specific information on the coding sequence and the full DNA sequence color coded by UTRs, exons, and introns. Throughout the instructor-led browser navigation, the instructor continually asks questions to underscore the students’ comprehension of the steps they are following, as opposed to merely showing the screenshots and listing instructions. It is helpful to have the instructor(s) and any instructional team members walk around the room and oversee the work of about twenty students each to ensure students follow along and are successful in finding the gene information. To bring the gene structure discussion back to the reasons we use this information in the lab, we have a brief conversation about where to design qPCR primers for *SOX2*.

Building on the instructor-led genome browser navigation completed together as a class, the instructor splits the class into small groups of ~4 students, who then work together on an in-class group activity to answer questions related to the structure of the *NANOG* gene, a gene which is more complex than *SOX2*. The in-class group activity exercise questions are on slide 30 of the included PowerPoint file (Supporting File S1), and the answer key can be found as hidden slides in the PowerPoint (Supporting File S1 [slides 31–34]) or as a printable document (Supporting File S2). The activity includes questions such as “draw the mRNA structure and determine the length of the transcript and coding sequence”, “locate the qPCR primers in the gene and mRNA transcript and determine the length of the amplicon”, etc. Students utilize portable whiteboards to draw the gene/transcript structure of *NANOG* and to answer the exercise questions together. Meanwhile, the instructional team walks around the room, answering student questions and probing students with additional questions to gauge their understanding. Some examples include: “How does the coding sequencing differ between the two *NANOG* transcripts? Are the qPCR primers located in an appropriate region of the gene, why or why not? What are the purposes of the 5′ and 3′ UTRs? Where does the ribosome load onto the mRNA and where does translation start?” etc. After the students complete the questions, they come back together as a class to present their gene structure drawings and to explain how they found the answers.

To give students the opportunity to practice using the databases more on their own, provide the students with the additional questions worksheet to complete outside of class (Supporting Files S3, S4.)

TEACHING DISCUSSION

We designed the lesson to focus on methods to allow students the opportunity to apply the concepts they learned during the lectures to a practical research lab environment. The 'How to Find a Gene' application activity followed lectures on gene structure, transcription, and translation, and preceded the qPCR and Western blot laboratory experiments. It aimed to explore gene databases such as NCBI and Ensembl, to learn how to navigate them, to find a detailed description about a gene of interest, to observe and appreciate the size of genes and their differing complexities, to get a sneak peek into primer designing for amplifying genetic material or quantifying cDNA, to predict protein sizes during protein purification or Western blotting, and to find other available biological information and data about genes.

The instructional team assisted during the active learning exercises, and students showed a high level of interest and enthusiasm in exploring both databases. A more extensive set of exercises that utilizes the UCSC genome browser to examine eukaryotic gene structure, transcriptional and translational units, and splice variants can be found in Laakso *et al.* (11).

Effectiveness of the lesson material

Many students were not familiar with these platforms and learned about the wide range of information available on genes and their products in many species. Overall, with the help of this lesson, students better understood the concepts of gene structure, transcription, and translation that they previously learned in the lectures and became comfortable with gene databases.

To facilitate learning, we used handouts with questions that allowed students to explore the NCBI and Ensembl databases. With the instructional team's help and collaboration with their peers, students were able to locate the necessary gene information and answered all questions correctly. Minimal instructor intervention was needed and most questions from students centered on locating appropriate *NANOG* gene information in the databases.

Students also navigated the databases to find protein information in a subsequent laboratory class where they performed a Western blotting experiment with blinded protein samples of different molecular weights. We provided them with a list of possible options for decoding the protein in their sample. After they obtained their experimental results, the students guessed their protein sample's molecular weight based on how it traveled on the gel in reference to the protein ladder, found the correct molecular weight for each option in the list using NCBI/Ensembl, and accordingly identified their sample.

A take-home exam was given following the lesson to measure student learning (Supporting Files S5, S6). The outstanding student performance on a take-home exam question dedicated to the NCBI/Ensembl databases suggested the lesson was effective. The question was of total 8 points. Scores ranged from 6.7 (81%) to 8 (100%) with a mean of 7.8 (97%).

Effectiveness of active learning exercises

Challenging students with questions during the refresher on central dogma, at the beginning of the section, helped students comprehend what would be taught later in class and fostered engagement and participation.

The instructor-led genome browser navigation using *SOX2* was used to guide the students through the two genome databases and find useful biological information. Visuals, such as screenshots of webpages, detailing what students should see on a particular page, and highlighting what students should click/select to find a piece of information, were particularly helpful to orient the students to the important information in the databases. These visuals prevented students from getting lost in the massive amount of information available on these websites and helped ensure they remained focused on the important content. PowerPoint slides and audio from the class were also made available after the section as a resource to refer to after class.

Students were required to work in groups of four to solve the in-class group activity. This collaborative learning helps students to learn from each other, increases communication and interaction between peers, increases inclusivity, and may help develop critical thinking skills (15). The instructional team observed that with collaborative brainstorming and each other's guidance, they answered most of the questions and required little help from the instructional team.

Student feedback on the lesson

After class and throughout the semester, students expressed their appreciation regarding this section. An anonymous mid-semester evaluation was conducted to garner students' feedback on the course components. Shown below are representative comments that we received for the question, "What is most effective about the exercise?"

"The NCBI and primer design courses were great."

"I like that we are learning about new databases on which we can analyze genes, make sequences, etc."

"The [exercises]...help apply the things we learn during lecture."

Additionally, we asked students to rate the effectiveness of the How to Find a Gene application exercise on a scale from 1 to 5, with 1 being very unhelpful and 5 being most helpful. Of the 42 total responses, the mean rating was 4.50 with 59% (25 students) rating it as very helpful, 33% (14 students) as slightly helpful, 4% (2 students) as neither helpful or unhelpful, and 2% (1 student) as slightly unhelpful.

Modifications

Collectively, we viewed the lesson as a success, but we recognize that it would be beneficial to dedicate more time to each feature of the database and each question on the in-class group activity. In our case, in-class time is restricted to 75 minutes, thus one strategy to overcome this challenge is to upload a pre-recorded video before the section that introduces students to the database's basic features. With this approach, we could briefly go over the basic functionality of NCBI/Ensembl and devote more time to the in-class group activity questions.

This lesson can be expanded for longer class periods or across multiple class sessions, as well as for more advanced students, by introducing and utilizing additional features of the databases. For example, Primer-BLAST, a primer designing tool from NCBI (16), can be introduced to design primers for the RT-qPCR experiment for measuring the expression levels of various genes during the differentiation from pluripotent stem cells to cardiomyocytes. Furthermore, for developing hypotheses and predicting experimental results, students can utilize tissue expression maps, such as Expression Atlas from Ensembl (17) or the Human Protein Atlas (18), to compare the gene expression levels in different tissues. For the human genes used in the experiment, students can also find information about orthologs in other model organisms and compare the expression levels across species. More advanced students can generate and upload a custom track of the students' gene models onto the UCSC genome browser and compare the Ensembl and NCBI transcript information in one view. Incorporation of such exercises would also provide effective ways to measure students' understanding of gene structure and database usage.

Due to time restrictions or apprehension, students may be unable to pose all the questions they might have during the allotted class time. A way to address this in the future is to offer open office hours (virtual or in-person) after the lesson. This additional time would allow the students to ask as many questions as they want, explore the database further, and create an opportunity for them to collaborate on the optional after-class handout.

While it was useful to walk through the database using screenshots of webpages while students performed the steps in real-time during the instructor-led genome browser navigation using *SOX2*, we found that students struggled to remember all the navigation steps when working through the in-class group activity on their own. Providing a handout detailing the steps presented in the exercise slides will help students when they are working by themselves. To this end, we have created an example handout to be used as a database reference sheet (see Supporting File S7). We plan to use this database reference sheet in our next implementation of the How to Find a Gene application activity. Additionally, NCBI has a series of how-to tutorial videos on YouTube that students may find helpful: A Beginner's Guide and Overview (this information is included in Supporting File S3).

Last, but certainly not least, this lesson presents a unique opportunity to integrate discussions on the sources and ethics of human genomics data into the classroom. Examples include discussing groups that have been used (and not used) for sequencing and annotating the human genome. This can be extended into a discussion of the more recent, inclusive human genome project that better captures the genetic variation present in different human populations (19) and how accounting for the complexities of human genetic variation can help to reduce racial bias in human populations (20). In addition, instructors can use examples that emphasize the importance of understanding single nucleotide polymorphisms present in certain groups, especially with respect to medical consequences. For example, after angioplasty, many patients are prescribed Plavix (clopidogrel), an antiplatelet drug used to prevent clotting, but patients from some Asian populations are resistant to this drug due to genetic polymorphisms (21). Integrating these discussions can enrich the lesson and help instructors to promote a more inclusive classroom.

SUPPORTING MATERIALS

- S1. How to Find a Gene – Lecture Presentation Slides and Exercise
- S2. How to Find a Gene – Exercise Answer Key
- S3. How to Find a Gene – Optional Additional Handout
- S4. How to Find a Gene – Optional Additional Handout Answer Key
- S5. How to Find a Gene – Take-Home Assessment
- S6. How to Find a Gene – Take-Home Assessment Answer Key
- S7. How to Find a Gene – Database Reference Sheet

ACKNOWLEDGMENTS

We would like to thank the instructional team (teaching fellows and undergraduate course assistants) who taught this exercise and the students who enrolled in the course during the spring 2020 semester for their feedback on the class activity. We thank Lisa Fountain, Mike Jin, and Saige Pompura for their valuable feedback on the manuscript. Finally, we are most grateful for the thoughtful editor and reviewer comments on our earlier draft, as they greatly enhanced our manuscript.

REFERENCES

1. Maloney M, Parker J, LeBlanc M, Woodard CT, Glackin M, Hanrahan M. 2010. Bioinformatics and the undergraduate curriculum. CBE Life Sci Educ 9:172–174. doi:10.1187/cbe.10-03-0038.
2. Banta LM, Crespi EJ, Nehm RH, Schwarz JA, Singer S, Manduca CA, Bush EC, Collins E, Constance CM, Dean D, Esteban D, Fox S, McDaris J, Paul CA, Quinan G, Raley-Susman KM, Smith ML, Wallace CS, Withers GS, Caporale L. 2012. Integrating genomics research throughout the undergraduate curriculum: A collection of inquiry-based genomics lab modules. CBE Life Sci Educ 11:203–208. doi:10.1187/cbe.11-12-0105.
3. Magana AJ, Taleyarkhan M, Alvarado DR, Kane M, Springer J, Clase K. 2017. A survey of scholarly literature describing the field of bioinformatics education and bioinformatics education research. CBE Life Sci Educ 13:607–623. doi:10.1187/cbe.13-10-0193.
4. Elgin SCR, Hauser C, Holzen TM, Jones C, Kleinschmit A, Leatherman J, Genomics Education Partnership. 2017. The GEP: Crowd-sourcing big data analysis with undergraduates. Trends Genet 33:81–85. doi:10.1016/j.tig.2016.11.004.
5. Lopatto D, Rosenwald AG, DiAngelo JR, Hark AT, Skerritt M, Wawersik M, Allen AK, Alvarez C, Anderson S, Arrigo C, Arsham A, Barnard D, Bazinet C, Bedard JEJ, Bose I, Braverman JM, Burg MG, Burgess RC, Croonquist P, Du C, Dubowsky S, Eisler H, Escobar MA, Foulk M, Furbee E, Giarla T, Glaser RL, Goodman AL, Gosser Y, Haberman A, Hauser C, Hays S, Howell CE, Jemc J, Johnson ML, Jones CJ, Kadlec L, Kagey JD, Keller KL, Kennell J, Key SCS, Kleinschmit AJ, Kleinschmit M, Kokan NP, Kopp OR, Laakso MM, Leatherman J, Long LJ, Manier M, Martinez-Cruzado JC, Matos LF, McClellan AJ, McNeil G, Merkhofer E, Mingo V, Mistry H, Mitchell E, Mortimer NT, Mukhopadhyay D, Myka JL, Nagengast A, Overvoorde P, Paetkau D, Paliulis

L, Parrish S, Preuss ML, Price JV, Pullen NA, Reinke C, Revie D, Robic S, Roecklein-Canfield JA, Rubin MR, Sadikot T, Sanford JS, Santisteban M, Saville K, Schroeder S, Shaffer CD, Sharif KA, Sklensky DE, Small C, Smith M, Smith S, Spokony R, Sreenivasan A, Stamm J, Sterne-Marr R, Teeter KC, Thackeray J, Thompson JS, Peters ST, Van Stry M, Velazquez-Ulloa N, Wolfe C, Youngblom J, Yowler B, Zhou L, Brennan J, Buhler J, Leung W, Reed LK, Elgin SCR. 2020. Facilitating growth through frustration: Using genomics research in a course-based undergraduate research experience. J Microbiol Biol Educ 21:40. doi:10.1128/jmbe.v21i1.2005.

6. Honts JE. 2017. Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. Cell Biol Educ 2:233–247. doi:1187/cbe.03-06-0026.
7. Grisham W, Schottler NA, Valli-Marill J, Beck L, Beatty J. 2010. Teaching bioinformatics and neuroinformatics by using free web-based tools. CBE Life Sci Educ 9:98–107. doi:10.1187/cbe.09-11-0079.
8. Sargent L, Liu Y, Leung W, Mortimer NT, Lopatto D, Goecks J, Elgin SCR. 2020. G-OnRamp: Generating genome browsers to facilitate undergraduate-driven collaborative genome annotation. PLoS Comput Biol 16:e1007863. doi:10.1371/journal.pcbi.1007863.
9. Hatfull GF. 2015. Innovations in undergraduate science education: Going viral. J Virol 89:8111–8113. doi:10.1128/JVI.03003-14.
10. Batut B, Hiltemann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, Bretaudeau A, Brillet-Guéguen L, Čech M, Chilton J, Clements D, Doppelt-Azeroual O, Erxleben A, Freeberg MA, Gladman S, Hoogstrate Y, Hotz H-R, Houwaart T, Jagtap P, Larivière D, Le Corguillé G, Manke T, Mareuil F, Ramírez F, Ryan D, Sigloch FC, Soranzo N, Wolff J, Videm P, Wolfien M, Wubuli A, Yusuf D, Galaxy Training Network, Taylor J, Backofen R, Nekrutenko A, Grüning B. 2018. Community-driven data analysis training for biology. Cell Syst 6:752–758.e1. doi:10.1016/j.cels.2018.05.012.
11. Laakso MM, Paliulis LV, Croonquist P, Derr B, Gracheva E, Hauser C, Howell C, Jones CJ, Kagey JD, Kennell J, Silver Key SC, Mistry H, Robic S, Sanford J, Santisteban M, Small C, Spokony R, Stamm J, Van Stry M, Leung W, Elgin SCR. 2017. An undergraduate bioinformatics curriculum that teaches eukaryotic gene structure. CourseSource 4. doi:10.24918/cs.2017.13.
12. Talbert R, Mor-Avi A. 2019. A space for learning: An analysis of research on active learning spaces. Heliyon 5:e02967. doi:10.1016/j.heliyon.2019.e02967.
13. Witt PL, Wheeless LR, Allen M. 2002. A meta‐analytical review of the relationship between teacher immediacy and student learning. Commun Monogr 71:184–207. doi:10.1080/036452042000228054.
14. Bengtsson L. 2019. Take-home exams in higher education: A systematic review. Educ Sci 9:267. doi:10.3390/edusci9040267.
15. Gokhale AA. 1995. Collaborative learning enhances critical thinking. J Teach Educ 7:22–30. doi:10.21061/jte.v7i1.a.2.
16. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 13:134. doi:10.1186/1471-2105-13-134.
17. Papatheodorou I, Moreno P, Manning J, Fuentes AM-P, George N, Fexova S, Fonseca NA, Füllgrabe A, Green M, Huang N, Huerta L, Iqbal H, Jianu M, Mohammed S, Zhao L, Jarnuczak AF, Jupp S, Marioni J, Meyer K, Petryszak R, Prada Medina CA, Talavera-López C, Teichmann S, Vizcaino JA, Brazma A. 2020. Expression Atlas update: From tissues to single cells. Nucleic Acids Res 48:D77–D83. doi:10.1093/nar/gkz947.
18. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA-K, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist P-H, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. 2015. Proteomics. Tissue-based map of the human proteome. Science. 347:1260419. doi:10.1126/science.1260419.
19. Khamsi R. 2022. A more-inclusive genome project aims to capture all of human diversity. Nature 603:378–381. doi:10.1038/d41586-022-00726-y.
20. Donovan BM, Semmens R, Keck P, Brimhall E, Busch KC, Weindling M, Duncan A, Stuhlsatz M, Bracey ZB, Bloom M, Kowalski S, Salazar B. 2019. Toward a more humane genetics education: learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations. Sci Educ 103:529–560. doi:10.1002/sce.21506.
21. Akkaif MA, Daud NAA, Sha'aban A, Ng ML, Abdul Kader MAS, Noor DAM, Ibrahim B. 2021. The role of genetic polymorphism and other factors on clopidogrel resistance (CR) in an Asian population with coronary heart disease (CHD). Molecules 26:1987. doi:10.3390/molecules26071987.

Table 1. Lesson Plan Timeline.

Activity	Description	Estimated Time	Notes
Preparation for class			
Prepare slides for instructor-led genome browser navigation and in-class group activity	Modify slides and prepare questions to be asked during class. If new genes are used, replace gene structure information and update screenshots from genome browser.	Varies depending on whether the same (*SOX2* and *NANOG*) or different genes are used for the activities.	• Lecture slides with notes are in Supporting File S1. • Questions to be asked during class may be added to lecture slides with notes. • Examples and questions in the exercise that are tied to a RT-qPCR laboratory experiment can be easily omitted from the rest of the lesson.
Class			
A refresher on central dogma and gene structure	Interactive lecture with active Q&A and diagrams of gene structure.	~10 minutes	• Lecture slides with notes are in Supporting File S1.
Instructor-led genome browser navigation	Interactive lecture with active Q&A and relevant parts pointed out on genome browser screenshots.	~20 minutes	• Lecture slides with notes are in Supporting File S1.
In-class group activity	Students work in small groups to answer questions that are projected on the slide.	~30 minutes	• Instructional team walks around the room to monitor group work and answer student questions. • Students may use whiteboards (or other means) to collectively draw and write their answers, to be shared with the rest of the class.
Presentations of in-class group activity	Group presents their answers to the entire class. Instructor explains answers to the questions and asks follow-up questions.	~15 minutes	• Encourage students to point to diagrams and answers they have drawn on the whiteboards. • Encourage students and instructional team to ask follow-up questions during group presentations. • Answers to these questions are found in Supporting File S1 (slides 31–34) and are also available as a printable document in Supporting File S2.
After class			
Optional after-class handout	Can be used as is or replace information and answer key for a new gene.	Varies	• Optional after class handout and answer key are provided as Supporting Files S3 and S4.
Take-home exam	Can be used as is or replace information and answer key for a new gene.	Varies; Students get ~1–1.5 weeks to work on the take-home exam, which contains questions from other lectures.	• Take-home exam questions and answers that are relevant to this lesson are provided in Supporting Files S5 and S6.