

Classroom as Genome: Using the Tools of Genomics and Bioinformatics to Illuminate Classroom Observation Data

Robert M. Erdmann^{1*} and Marilyne Stains^{1*}

Department of Chemistry, University of Nebraska—Lincoln, Lincoln, NE 68588

ABSTRACT

Classroom observation protocols can provide an exceedingly rich form of data. However, this is a double-edged sword, as researchers often struggle to take full advantage of the detailed data outputs. In this essay, we introduce a new approach to the analysis of classroom observation data, termed “classroom as genome” (CAG). We illustrate how real-time classroom observation data and genomic data can be viewed as quite analogous, both conceptually and in terms of downstream analysis. We provide both abstract and concrete examples of how the tools of genomics and bioinformatics can be applied to classroom observation outputs. We also show how this philosophy of analysis allows for the layering of information from multiple observation protocols onto the same classroom data. The CAG approach enables biology education researchers to explore detailed patterns within observed classrooms in a highly scalable manner.

INTRODUCTION

We need to think creatively about how to bring life sciences research methods—such as those used to study physiological systems, to model ecological processes across scales, and to analyze metabolic networks—to bear on the study of teaching and learning.

Dolan, 2015, p. 1

Researchers and instructors can use a variety of tools when seeking to better understand classroom practices, including surveys, classroom observation protocols, and interviews (American Association for the Advancement of Science, 2012). While each mechanism for data collection has pros and cons, classroom observation data have been highlighted as potentially more objective when compared with surveys and interviews relying solely on instructor reports (Ebert-May *et al.*, 2011). A growing number of classroom observation protocols have been developed to convert video and audio recordings (or the observations of the trained eyes and ears of live observers) into more easily interpretable and comparable data structures (Supplemental Material). While the focal point of each observation protocol may differ, they all share the goal of reflecting the reality of the classroom experience as accurately as possible.

Classroom observation data are useful in a range of contexts. For individual instructors looking to better understand and improve their own classroom practices, classroom observations may provide a variety of insights (e.g., Achen and Lumpkin, 2015). For education researchers, observing and analyzing classes at larger scales can shed light on the instructional practices present within a broader population. For instance, the Course Observation Protocol for Undergraduate STEM (COPUS; Smith *et al.*, 2013) has been used to observe classes at a variety of postsecondary institutions. Resulting data have been used to characterize the wide spectrum of instructional practices found within a single research university (Smith *et al.*, 2014), as well as to

Ross Nehm, *Monitoring Editor*

Submitted Jul 16, 2018; Revised Dec 6, 2018;
Accepted Dec 17, 2018

CBE Life Sci Educ March 1, 2019 18:es1

DOI:10.1187/cbe.18-07-0116

¹Present address: Center for Learning Innovation,
University of Minnesota Rochester, Rochester, MN
55904.

*Address correspondence to: Robert M. Erdmann
(rerdmann@r.umn.edu) or Marilyne Stains
(mstains2@unl.edu).

© 2019 R. M. Erdmann and M. Stains. CBE—Life
Sciences Education © 2019 The American Society
for Cell Biology. This article is distributed by The
American Society for Cell Biology under license
from the author(s). It is available to the public
under an Attribution–Noncommercial–Share
Alike 3.0 Unported Creative Commons License
([http://creativecommons.org/licenses/
by-nc-sa/3.0](http://creativecommons.org/licenses/by-nc-sa/3.0)).

“ASCB®” and “The American Society for Cell
Biology®” are registered trademarks of The
American Society for Cell Biology.

compare, contrast, and cluster hundreds or even thousands of class periods from a range of North American classrooms (Lund *et al.*, 2015; Stains *et al.*, 2018).

Analyzing classroom observation data at large scales is not without difficulty, however. For instance, Smith and colleagues remarked that “it was difficult to get a general sense of trends in student and instructor behavior when comparing 25 possible COPUS codes in 51 different courses” (2014, p. 627). In a way, this makes these observations one of education research’s many potential forms of “big data”—data that cannot be easily analyzed through conventional or straightforward analysis methods (Madhavan and Richey, 2016). Large-scale studies such as those referenced above frequently rely on the use of summary statistics to compress and restructure the data before conducting downstream analyses. However, the use of summary statistics alone has the potential to obscure interesting trends and meaningful differences. This phenomenon is epitomized in Anscombe’s quartet, a group of four data sets that are clearly quite different when viewed in graphical form (e.g., a parabolic arc is by eye entirely distinct from a straight line), but are nearly identical in terms of descriptive statistics (Anscombe, 1973). A challenge for the field is to complement current big-picture, summary statistical analysis techniques with new methodologies that allow researchers to fully explore even the most fine-grained versions of these enormous data sources.

In this essay, we wish to introduce a new philosophy of classroom observation data structuring and analysis—the “classroom as genome” (CAG) approach. We draw upon the tools and

mindsets of the bioinformatics scientist, with the goal of repurposing genomics analysis methods to enable new approaches to specific research questions within science education. We argue that the CAG approach will allow the biology education research community to delve far deeper into classroom observation analysis, both at the single-classroom level and at highly parallel scales with the potential to analyze hundreds of classes and instructors simultaneously, without a concurrent need for the development from scratch of new tools or software.

CLASSROOM AND GENOME: PARALLEL CONSTRUCTS

The similarities and differences of classroom and genome within our analogy are summarized in Table 1.

Layers of Information

The genome of a species can be summarized as a sequence of DNA bases strung in a long line, but in the hands of a genomics researcher, that is merely the scaffold upon which all manner of genomic information and analyses can rest (Figure 1A). A more comprehensive picture of a genome incorporates many layers of information, such as the annotations of transcribed mRNAs, the DNA methylation status of cytosine bases, the prevalence of all manner of chromatin modifications, and the abundance of small RNAs derived from the underlying sequence. These genomic features described are generally not mutually exclusive—at any given point along the DNA sequence, multiple types of overlapping genomic features can co-occur.

TABLE 1. The classroom as genome analogy: similarities and differences

	Concept	Genomic example	Classroom example
Similarities	Ability to layer information of various types	DNA methylation, mRNA transcripts, and chromatin can all be viewed or analyzed simultaneously.	Student questions, instructor answers, and formative assessments can all be viewed or analyzed simultaneously.
	Genome as a defined scaffold	The location of a promoter or a gene can be easily described as a chromosomal position.	The timing of an instructor behavior or a student response can be easily described relative to the start or endpoint of a class period.
	The chromosome as an independent unit of analysis	The contents of chromosome 1 are often analyzed independently of, but in parallel to, the contents of chromosome 2.	An individual class period can be analyzed independently of, but in parallel to, a second class period.
	Interactions between genomic elements	Chromosome conformation capture illustrates that certain chromosomal regions can preferentially interact with specific regions of other chromosomes.	Each class period within a broader course may be an independent entity, but there are generally threads of related content or repeated activities that bind these independent units.
	Patterns and directionality matter within a genome	The positioning of a promoter upstream of a gene is often essential for proper transcription—downstream positioning of the same genetic element would not have the same outcome.	The positioning of certain classroom activities relative to one another in time has important impacts on the learning environment—a student question before an activity may infer something very different than the same student question following the same activity.
Differences	Universality of directionality	With sense and antisense strands, transcription occurs bidirectionally.	Classroom time only runs from start to finish.
	Definition of a genome	For the genome of a given species, scientific community comes to a consensus on numbering/labeling for chromosomes.	“Genome” definition is completely free-form and dynamic depending on data to be analyzed.

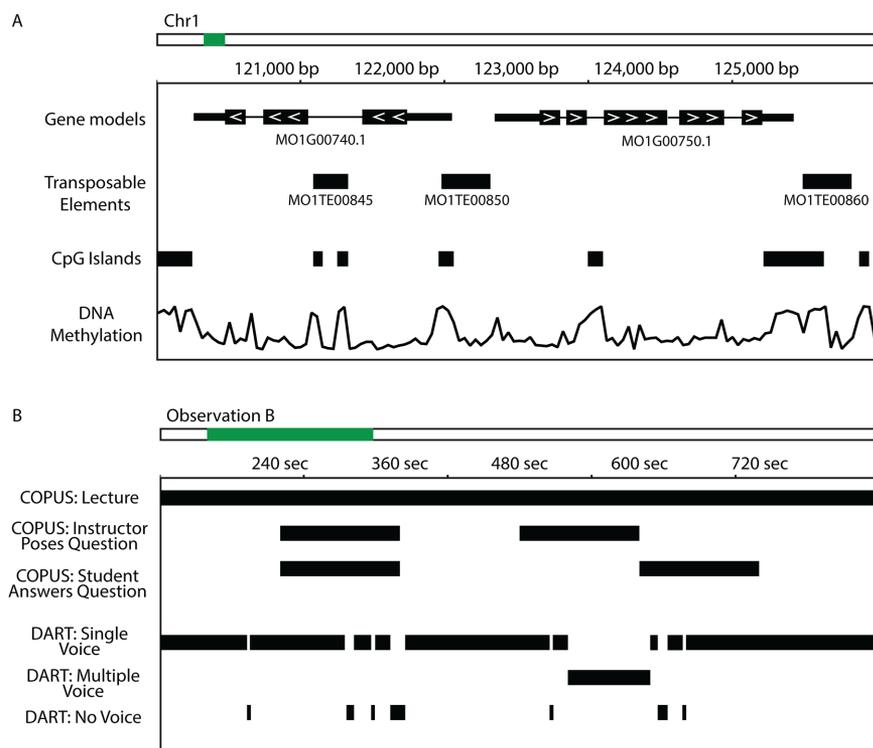


FIGURE 1. Mock genome browser and classroom browser displays. (A) Sample genome browser display, illustrating a portion of chromosome 1 for a model organism. The area in green above the browser window depicts the portion of the chromosome in view. (B) Sample classroom browser display, illustrating a portion of a classroom observation with select COPUS and DART codes (see the Supplemental Material) visible.

Classroom observation data can be thought about in much the same way. Information from multiple classroom observation protocols can be layered onto the same class period, with the timeline of the classroom serving as the scaffold to which everything maps (Figure 1B). As in the case of biological data, a variety of codes¹ might be overlapping at any given position within a hypothetical classroom genome. By retaining the connection to a common yardstick for each code, rather than merely reducing data to median/mean values, we are able to glean information about different behavior types overlapping or flanking each other, instead of thinking of them as independent entities that cannot be directly related to one another.

One key note is that, even as we describe the power of overlapping data type over data type onto a single scaffold, one trait illustrated by a range of bioinformatics studies is that access does not equate to mandate. A chromatin researcher might be much more interested in contrasting all of the histone markers associated with a region of the genome and might find little value examining small RNA expression patterns at the same time, even if that information were readily available. Similarly, in an education research capacity, researchers do not have to use all available code types all of the time—for example, an analysis of instructor behavior might omit data relating to student behaviors for the sake of simplicity and coherence.

¹Throughout this essay, we will use the word “code” to describe any data classification or annotation type that is associated with an observation protocol.

Directionality and Genomic Superstructure

Directionality has meaning for both biological genomes and classroom “genomes.” In the case of biological chromosomes, the labeling of one end of a chromosome as “base 1” and the other end as “base 20,000,000” is arbitrary, but because genes have meaningful starts and ends, directionality is encoded into the system. A genetic sequence might exert a completely different effect on a gene’s expression if it followed the gene instead of preceding the gene. In the case of a classroom, this is also the case—actions taken by instructors and students are contextualized by what events precede or follow them. Furthermore, the scaffold of time that the classroom genome is built upon has a very easily understood directionality—time only marches forward.

The split of genomes into separate chromosomes, each containing a subset of the total DNA complement, provides the genomes with a superstructure. These chromosomes are independent and bounded entities (e.g., the start of chromosome 2 is generally not immediately adjacent to the end of chromosome 1). Consequently, certain genomic-level analyses are conducted separately in parallel for each individual chromosome, even if the entire genome is under study. In our

conception of classroom observation data, each class period is analogous to an individual chromosome. Even if there are videos of multiple classes taught by the same instructor, each class should be considered as an independent unit, especially in light of the fact that any individual class taught by an instructor should not be considered fully representative of that instructor’s pedagogical practice (Lund *et al.*, 2015; Stains *et al.*, 2018).

If each class period is an individual chromosome, what is the collection of chromosomes comprising the “genome” in this analogy? Here, we highlight a difference between the biology and education versions: in the case of our classroom example, class period “chromosomes” might not be mutually exclusive to a single genome. One analysis might place class period “A” into an “Instructor #1, Biology 101 course” genome, perhaps along with class periods “B” and “C” that were observed from the same course. Another analysis might place class period “A” into a “Socratic instruction” genome, along with class periods from other instructors and courses, while placing “B” and “C” into a “group-work” genome. So long as the chromosomes remain as independent units of analysis, the conception of the overarching genome as an organizational tool remains completely flexible.²

²This flexibility does require that researchers adopt internally consistent naming conventions for chromosomes and files. While this is clearly a best practice for data analyses of any type, it takes on additional importance in a situation in which there is no universally accepted set of known chromosomes to work from, as is the case in most biological genomics analyses.

While chromosomes are a logical unit for independent analysis, different chromosomes do interact in interesting and important ways. For example, genes in the same biological pathway can be located on separate chromosomes, and an entire suite of biological techniques has been developed to connect linear genomic representations with three-dimensional closeness metrics within the nucleus (Dekker *et al.*, 2002). Similarly, in the classroom, individual class periods generally happen within the context of a larger course. An activity that happens only once in an observation of a class period may also have happened in a dozen preceding class periods or may never have happened before in that course. If the data allow it, analyzing course structure at the same time as class period structure allows researchers to see both the forest and the trees.

Underlying Patterns

In many ways, the field of genomics is first and foremost the study of patterns and associations between different forms of biological information. Nature provides us with many interesting examples that dovetail with our understanding of biological processes at the molecular level. Various forms of promoter sequences precede genes, helping to direct the necessary transcriptional machinery to an appropriate starting point. In general, active chromatin marks show a strong association with highly transcribed genes. Zooming out, large regions of a chromosome may share a broad chromatin signature. Taken together, these patterns and associations make it evident that genomic features frequently share relationships with other features, with clearly understood biological consequences.

Classroom observations have the ability to illuminate similar types of patterns and sequences. For example, if an instructor poses a nonrhetorical question to the class, it is likely that students will provide some type of answer in response. (Just as in a biological scenario, there are always exceptions, and these are often every bit as interesting as the majority patterns.) Furthermore, any such student responses are likely to be reacted to in some way by the instructor—through acceptance, clarification, a follow-up question, or redirection, among other options. In this scenario, certain codes are more tightly coupled than others, but the potential for underlying sequencing is quite evident. This means that genomics approaches for finding meaning within large, complex data sets should also provide ways to pursue the analysis of patterns and configurations embedded in instructional events.

THE POTENTIAL TO ADDRESS WIDE-RANGING RESEARCH QUESTIONS

Analysis solely for the sake of analysis does not advance the research enterprise—the CAG approach is useful only insofar as it enables researchers to address meaningful research questions. Here, we offer a spectrum of sample research questions, directly followed by the more “technical” questions that underpin the answering of the research questions. These technical questions are the actual questions addressed by CAG methods.

How Long Is Instructor Wait Time after Posing a Question to the Class? → What Is the Mean/Median Length of a Single Instance of a Particular Code?

When behaviors occur, are they brief or extended in nature? The genomics tools used to enable CAG can easily determine

this, similar to calculating the mean lengths of genes or other features found on a particular chromosome. For example, a researcher interested in instructor wait time following questions might encode data using Flanders interaction analysis (FIA; Amidon and Flanders, 1967), which includes code 10 (Silence) and code 4 (Instructor poses a nonrhetorical question). Then, the researcher could determine the average span of an unbroken stretch of code 10 immediately following an instance of code 4 across any number of classes of interest.

How Evenly Dispersed Are Clicker Questions across Entire Periods of Instruction? → How Far Is This Code from Other Instances of the Same Code?

These questions get at the dispersion of a behavior over the course of a class—scattered versus tightly clustered. If 20% of a class period is occupied by a particular behavior/code, it could make a big difference if that 20% occurs in a single extended instance or is dispersed as multiple smaller instances widely spread out across the class. In this case, a researcher could examine the distribution of clicker question annotations relative to their nearest neighbors to better describe the clustering (or lack thereof) of the code.

Are Student Questions and Instructor Questions Internally Segregated/Clustered within a Period of Instruction or More Evenly Intermixed? → How Far Is This Code from Instances of Different Codes?

A deeper understanding of classroom events often requires the analysis of relationships between different codes. For example, a researcher might want to examine the relationship between student- and instructor-posed questions. To do this, relative distance metrics can be employed to show how the two behaviors are situated relative to one another over a class period, as will be illustrated in a case study later in the essay. Determining the dispersion and intermingling of multiple codes as opposed to a single code allows for the identification of broader pockets of associated activities that extend far beyond the start and end of any particular classroom event.

When an Instructor Is Not Actively Doing Anything, What Else Is Happening in the Classroom at That Point? → What Codes Overlap with This Code? When This Code Is Compared with Other Codes, How Similar Is the Coverage?

By determining which codes overlap with a code of interest, a researcher has the ability to assess correlations between different behaviors. For example, the COPUS code IW (Instructor waiting) might show considerable overlap with certain other codes, and determining such correlations might provide a window into how instructors conceive of associated (or nonassociated) classroom activities. This is useful even within the context of a single observation protocol, if codes are nonmutually exclusive within a time block. However, this ability becomes particularly interesting when more than one observation protocol is overlaid on the same class period. When this happens, relationships between codes from different protocols can be observed, expanding the pool of potential broader questions that can be answered.

What Codes Precede Instructor Questions? → What Happens before (or after) This Code?

In this case, a researcher may want to see what activities or behaviors are enriched during the lead-up to questions posed to the class. The usage of ends analysis, described further in the case studies to follow, enables a researcher to compile a detailed and complete picture of the regions flanking a code of interest. Addressing such questions is often interesting in its own right, but identifying the codes preceding or trailing another code is the fundamental basis for a variety of potential analyses regarding the sequencing and patterning of codes.

What Is the Sequencing of Codes That Surrounds a Clicker Question? → Is This Code Part of an Overarching Pattern/Sequence?

Taking many of the prior questions into account, a researcher can begin to ask questions addressing overarching patterns. Within any given field of instruction, instructional style, class demographic profile, or other categorization of interest, common instructional signatures and patterns may exist that cannot be easily observed through other means. CAG approaches enable this sort of higher-order observational analysis and may spur new avenues of inquiry. For example, the instructional patterns associated with the usage of clicker questions might inform a study on the fidelity of implementation of peer instruction. Ultimately, this level of question should enable a better examination of the linkages between classroom activities and student outcomes—a fundamental goal of much of the field's research.

In most cases, researchers seeking to address these types of questions must have large samples for the signal to emerge from the noisiness inherent in this type of data. If a researcher is examining a single classroom observation, some of these technical questions might be answerable without inordinate difficulty, but the use of CAG enables the grouped assessment of hundreds of class periods simultaneously, a prospect inconceivable through manual approaches.

CAG CASE STUDY ANALYSES

To illustrate how such analyses might work in an actual education research context, we will present a number of case study examples, using COPUS data from a prior publication (Lund *et al.*, 2015) as the starting point for further analysis. These data were reformatted into a genomics file format before analysis using publicly available bioinformatics tools and workflows (see *Methodology for Implementation* section and the Supplemental Material for details).

The Context of the Clicker Question Code

Imagine that you are a researcher who is very interested in the ways in which clicker questions are implemented within observed classrooms. The COPUS code Clicker question (CQ) would likely be of strong relevance to your research. As discussed in the preceding section, there are many things you might be eager to explore about classroom time blocks in which CQ was coded. Here, hypothesis testing can be easily envisioned—perhaps you hypothesize that when instructors use clickers (CQ), they are more likely to ask students to talk to one another before voting (CG) than to have students answer the clicker questions individually (Ind).

This type of question can be addressed through the production of an ends analysis plot, a tool used by genomics researchers to visualize features that precede, follow, or overlap other features of interest—for example, where transposable elements tend to sit relative to genes in the genome. In ends analysis, the starting end and the closing end of every instance of a code/feature of interest are lined up, while all other codes are plotted relative to those aligned ends.³ In Figure 2, we have produced an ends analysis plot that shows how select COPUS codes relate to the instances of CQ codes found within 269 classroom observation periods (only the 10 COPUS codes with the highest average frequency around CQ codes are displayed).

Figure 2 indicates that our hypothesis holds true: we see that, even though both CG and Ind frequencies increase drastically after a clicker question has been asked, Ind climbs only to a maximum frequency of 34%, compared with CG, which climbs to a frequency of 83%. However, the distribution of the two codes complement one another, with Ind coming to its highest value near the starting point of the CQ codes, while CG starts at a below-peak value of 59% and then increases during the later phases of the CQ codes.⁴ This analysis illustrates how CAG provides opportunities to move beyond looking at an individual code and relate the patterning of multiple codes simultaneously.

If we shift to a more exploratory mode of analysis, what broader patterns can be discerned? We find that listening, and to a lesser extent, lecturing, are the codes most often found preceding the CQ code locations. When we shift our focus to the overlapping CQ area in the center of the plot, we observe a strong spike in the Instructor waiting (IW) code, reaching a peak of 86% overlap with CQ. Instructors within the sample were an order of magnitude more likely to wait during a clicker question than to move through the classroom (MG code—not shown in Figure 2 due to low frequency). Meanwhile, the Follow-up (Fup) code spikes near the end and immediately following the CQ code, showing that the end of a clicker question is often followed by some form of follow-up. While here we have focused on only a small number of high-frequency codes, we have already started to compile a picture of how CQ may correlate with other codes among our observed population.

Such an analysis is a simple example of the type of pattern discovery that can be done through CAG approaches. This is only the first layer of potentially many. For example, we found that clicker questions had an overlapping individual thinking (Ind)

³Just as genes can have a variety of lengths, classroom activities can take varying amounts of time. If one simply lines up the starting points of all features, the ending points would not align, and vice versa. To adjust for this, we can construct ends analysis plots in one of two ways: as a continuous plot or as a split plot with a vertical gap in the middle. To make a continuous plot, we must transform the data from a time-coupled form into a fraction-of-overall-length form. In a split plot, like the one shown in Figure 2, the starting points and the ending points are aligned, and a certain amount of overlap is displayed, but any features longer than a certain length will not be fully characterized in the center region.

⁴In some ways, this description is a slight simplification of a complex scenario. In COPUS data, Ind could easily overlap with CG in a peer-instruction scenario in which students first think individually, then discuss with their neighbors. Thus, CG numbers cannot be taken to only mean “students jump straight into discussion with peers without individually considering the question.” However, analysis of the frequency with which CG codes overlap or follow Ind codes shows that our description in the main text is a fair characterization of the practices within our sample.

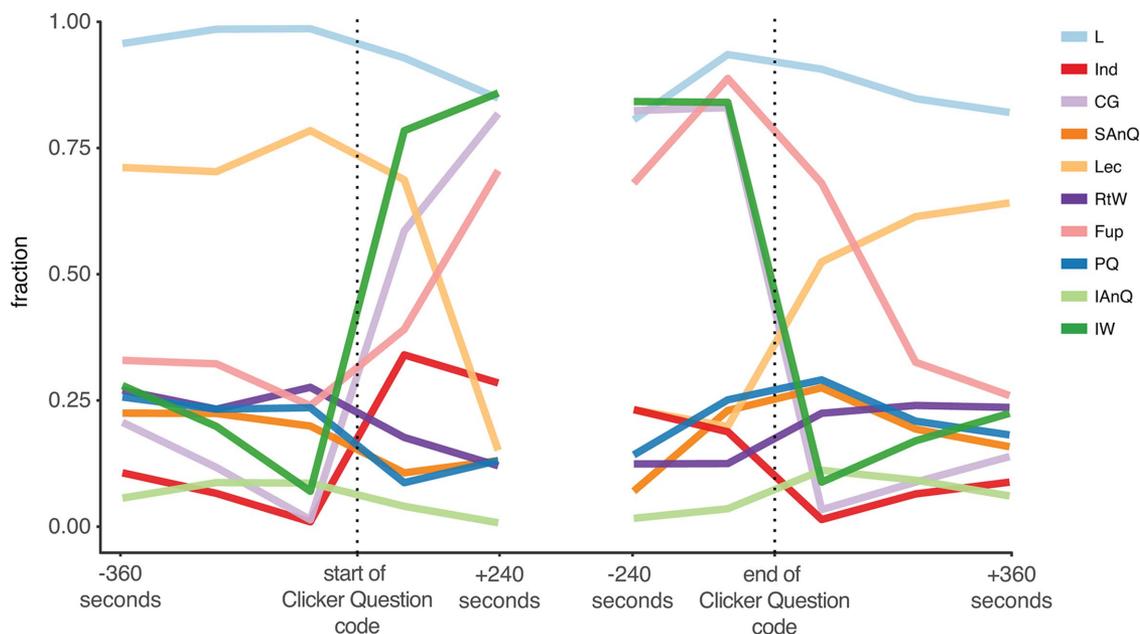


FIGURE 2. Ends analysis showing the relationship of select COPUS codes with a compilation of all CQ code occurrences found within 269 classroom observations. Additional details regarding ends analysis plot construction and interpretation can be found in the Supplemental Material. Abbreviations: L, Students listening; Ind, Student independent work; CG, Group discussion of clicker questions; SAnQ, Students answer questions; Lec, Instructor lectures; RtW, Instructor writes on board during instruction; Fup, Follow-up to clicker question or activity; PQ, Instructor poses question; IAnQ, Instructor answers question; IW, Instructor waits.

component a little more than a third of the time in our sample. However, Ind might be even more tightly associated with CQ than this suggests. Indeed, we find that for greater than 70% of Ind code instances, a clicker question is asked simultaneously. Clearly, analyses need to account for the asymmetric nature of code overlap to provide a complete story. In this way, we can begin to spur on new, unforeseen research questions as well as provide better descriptions of general patterns and subpatterns of classroom activity.

The Layering of Classroom Observation Data

One of the most valuable features of the CAG approach is the fact that different observation protocols can be applied to the same classroom data. This works because CAG is fully instrument agnostic—researchers can use whatever observation protocols address their research questions of interest and overlap complementary types of analysis. Thus, codes and components can be mixed and matched in whatever way advances a study.

As an example, imagine that you are a researcher who is interested in the way in which instructors and students ask and answer questions within the classroom. More than one observation protocol could be useful in this regard. The previously mentioned COPUS contains a number of codes corresponding to student and instructor questions. Meanwhile, FIA categorizes instructor and student verbal interactions, including questions and responses, into 10 codes (Amidon and Flanders, 1967). These two observation protocols differ in several respects. For example, the quanta of COPUS codes are much larger (120-second blocks) when compared with FIA codes (3-second blocks). This means that using COPUS may lead to the loss of information—a classroom event that takes 5 seconds may be coded similarly to something that

takes 50 seconds. Conversely, FIA only has 10 associated codes, and this can lead to a loss of information when compared with the 25 codes of COPUS. These differences may result in the two protocols providing complementary information when applied to the same classroom observation, potentially mitigating the drawbacks of each, and opening up new avenues of analysis.

To illustrate this type of dual application of observation protocols, we coded a subset of our COPUS-coded classroom videos using FIA ($N = 5$). The videos selected for FIA coding were classified as “Socratic—at board” in Lund *et al.*, 2015. (Details of the analysis approach can be found in the Supplemental Material.) If we compare the percentage of class time occupied by codes that represent similar constructs, we can see how the two protocols might be complementary in terms of deeper analysis. In Table 2, we show five classroom observations, each represented by a row, and four classroom observation protocol codes, each represented by a column. Each percentage reported represents the percentage of the total time from an individual classroom

TABLE 2. Comparison of selected COPUS and FIA codes for “Socratic—at board” cluster class observations

Observation	Instructor poses question (COPUS)	Instructor asks questions (FIA)	Student question (COPUS)	Student talk—initiation (FIA)
1	48.0%	7.9%	0.0%	0.0%
2	48.6%	2.6%	43.2%	3.7%
3	42.9%	5.5%	40.0%	6.8%
4	68.0%	8.1%	16.0%	0.9%
5	50.0%	6.0%	3.8%	0.8%

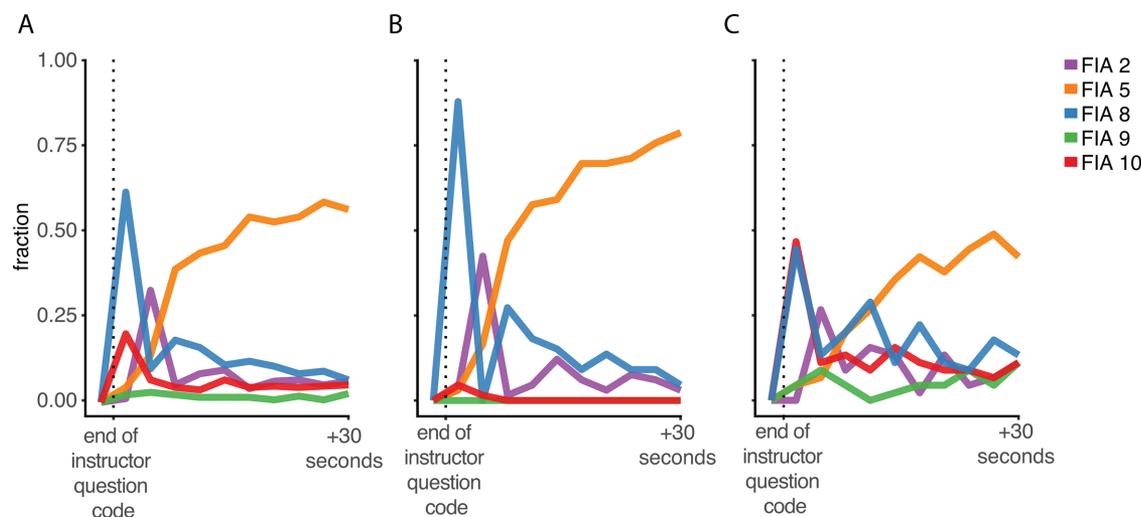


FIGURE 3. Ends analysis showing the frequency of five FIA codes in the period following the end of the FIA 4 code (instructor asking a nonrhetorical question). (A) Compiled profile for five classroom observations drawn from the “Socratic—at board” COPUS profile (Lund *et al.*, 2015). (B) Profile for observation 1 in Table 2. (C) Profile for observation 3 in Table 2. Abbreviations: FIA 2, Instructor praises or encourages student; FIA 5, Instructor lectures; FIA 8, Student talk in response to a teacher’s question or instructions; FIA 9, Student talk initiated by the student; FIA 10, Silence.

observation that was coded for the code associated with the column. Here, we can split the four codes into two meaningful pairings of codes: COPUS PQ (Instructor poses a question) pairs with FIA 4 (Instructor asks questions), and COPUS SQ (Student asks a question) pairs with FIA 9 (Student talk that is initiated by the student). In the case of the first pairing of codes (the “Instructor” columns of Table 2), observations 1 and 2 are of particular interest. While the COPUS percentages are quite comparable, the FIA percentages are divergent, with the instructor of observation 1 apparently spending roughly three times as much time posing questions to students according to FIA. On the other hand, in observations 4 and 5, the FIA code for students initiating talk/posing questions shows very comparable values, but the percentage for the student question COPUS code shows a four-fold difference between the two classes, implying a difference in student question dispersion. As the table suggests, any observation protocol may have shortcomings or oversimplify things in certain situations, but through the usage of additional protocols on the same underlying data, some of these situations can be better identified and mitigated.

The first way in which we used the complementary information of the two observation protocols was by leveraging information from one protocol to select a sample for the application of the other. When clustered by select COPUS code proportions (Lund *et al.*, 2015), the selected class videos all fell into the same bin and should thus be representative of a certain type of class structure (in this case, lecturing with Socratic questioning while writing on the board). As described earlier, this is one way (of many) to organize a coherent “genome” of classes to analyze in tandem. Once these classes were subjected to FIA coding, we could then perform analyses similar to those described earlier for the COPUS CQ code. As an example, we could examine FIA code 4, which corresponds to an instructor posing a nonhypothetical question to his/her students. In Figure 3A, we provide an ends analysis that illustrates, on average, what FIA codes follow FIA 4 in our sample of analyzed “Socratic—at board” observations.

In these classes, a student response to the instructor question (FIA 8) is the most common code within the 3 seconds immediately following the end of the instructor question, occurring roughly two-thirds of the time, with Silence (FIA 10) representing the second most common code (most often corresponding with wait time). Between 3 and 6 seconds after the end of the question, the two codes just mentioned become much less common, and instructor praise or acknowledgment of student answers becomes the most common code (albeit only with a plurality). By 9 seconds after the end of the question, the lecture code becomes the most common code, and continues to be so for an extended stretch of time. Therefore, on average in this sample of five observations, it takes about 9 seconds for students to process, respond to, and be recognized for their answers to an instructor’s question.

Using a global analysis approach does not preclude the discovery of individual differences within our sample. In Figure 3, B and C, we illustrate ends analysis performed on two individual class periods within the FIA-analyzed sample of “Socratic—at board” observations, as opposed to the overall average. Here, we observe a stark difference in the pattern of codes following the instructor’s questions. In Figure 3B, nearly 90% of the time following an instructor question, the first code observed was a student response to the question. In Figure 3C, student responses to the question were not even the most common code following the end of the instructor’s question—instead, silence was the most common occurrence. Looking at the overall code patterns, a clear difference in the use of wait time after the question is apparent. By using FIA in addition to COPUS, we are able to highlight differences in classroom dynamics that COPUS alone would not highlight, and these differences could easily lead to further investigation.

There are other ways in which an observation protocol pairing could be used. In genomics, relative distance is a metric that describes the spatial correlation between a pair of feature types (Favorov *et al.*, 2012). For example, if an

instance of code 2 overlapped with code 1, the relative distance metric between code 1 and code 2 at that position would be 0, while if another instance of code 2 was exactly equidistant between the nearest flanking instances of code 1, the relative distance metric would be 0.5 (Figure 4A). With a larger number of features, a frequency distribution can be calculated (Figure 4B). A uniform or flat distribution of frequencies suggests random assortment of the codes, while an above-average frequency at low relative distance would suggest that the codes tend to be closer together than would be expected from a random assortment.

If we wanted to probe the relationship between select COPUS and FIA codes, we could use relative distance as a measure of the level of spatial association between the codes in question. In Figure 4, C–F, each relative distance plot summarizes the relationships between select FIA codes and a single COPUS code. The slopes of the linear regressions provide an indication of the relationships between the codes over the totality of the class periods analyzed, with negative slopes indicating that codes are closer together than expected by chance, positive slopes indicating that codes are farther apart than expected by chance, and flat slopes indicating a relatively random distribution relative to the other code. In Figure 4D, we observe that the COPUS code SQ (Student asks question) tends to be much closer to instances of FIA 9 (Student talk initiated by the student) and FIA 3 (Instructor accepts or uses ideas of students) than expected. The former association is not surprising, simply due to the closely paralleling definitions of the two codes, but the latter association is potentially more interesting, suggesting that instructors are much more likely to accept or use the ideas of students close in time with student questions.

Figure 4 also alludes to potential avenues for further investigation. Figure 4, C and E, relates to the COPUS codes PQ (Instructor posing questions) and SQ (Students answering questions), respectively. In both cases, FIA 4 (Instructor asks nonrhetorical questions) and FIA 8 (Student talk in response to a teacher's question or instructions) have negative slopes, indicating a logical close proximity of these FIA codes to the COPUS codes. Meanwhile, the FIA 4 and 8 codes have positive slopes in Figure 4, D and F, which features the COPUS codes SQ and IAnQ (Instructor answers a question). This implies that instructor-initiated question events are more distant temporally from student-initiated question events than expected by chance. Altogether, this may indicate that during Socratic—at board instruction, for both instructor and student questions, it is more likely that the nearest neighboring questions within a class period would be questions of the same type, as opposed to alternating between question types. This is an example of how CAG analyses can illuminate potential patterns of instructional practices at large scales.

We can display the magnitude of data at our disposal, representing both COPUS and FIA coding at the same time, in a genomic-style representation (Figure 5). This figure displays an overwhelming amount of information, with thousands of data points. However, the point of using genomic tools for analysis is to condense this mass of data into digestible observations and patterns that can contribute to our understanding of classroom behaviors. These observation protocol outputs, as complex as they are, can be easily transformed into anything from basic summary statistics to

complex ends analysis plots for codes of interest. At its heart, the idea of CAG is to look at something like Figure 5 and see vast untapped potential and possibilities, enabled by tools already at a researcher's disposal, as opposed to something inscrutable, beyond any individual's comprehension or understanding.

METHODOLOGY FOR IMPLEMENTATION

Now that we have discussed the purpose and potential of CAG approaches, we can more easily delve into potential methods for implementation. In this section, we focus on conveying the broad theory and basic mechanics, with greater detail available in the Supplemental Material.

Translating Classroom Observations into Genomic File Formats

How do classroom observation data get converted into a “genomic” form?⁵ There are multiple file types used in bioinformatics analyses, but here, we will use bed (browser extensible data) files as a representative example. The bed file format, at its heart, is simply a tab-delimited notation of the chromosome number, starting position, and ending position of each feature, listed sequentially. For example, we might indicate the positioning of four members of a gene family of interest within a genome as such:

```
Chr1 100 1600
Chr1 2000 3200
Chr2 6000 8000
Chr4 2500 3600
```

By convention, the file describes the presence of something at the specified location, like mRNA transcripts of interest (or in the case of CAG, perhaps an instructor lecturing to the class). However, it is certainly possible to create bed files that denote the absence of a particular code or codes—for example, areas of the genome that *do not* overlap with annotated genes or transposable elements or times during a class period when the instructor is *not* lecturing.

An optional fourth column within the bed file format provides a space for the inclusion of additional information. In many cases, biological data associated with a location will not fit a strict binary, and this fourth column provides a mechanism for retaining this information in addition to positioning data. For example, we could show the average level of CpG methylation within the specified genomic spans as such:

```
Chr1 100 200 0.67
Chr1 2000 2200 0.50
Chr2 7800 8000 0.20
Chr4 2550 2650 0.87
```

⁵This conversion can be done manually, but is best done through automation, not only to save time, but also for accuracy and reproducibility reasons. Automation looks different based on the input file type or format and is discussed in greater detail in the Supplemental Material.

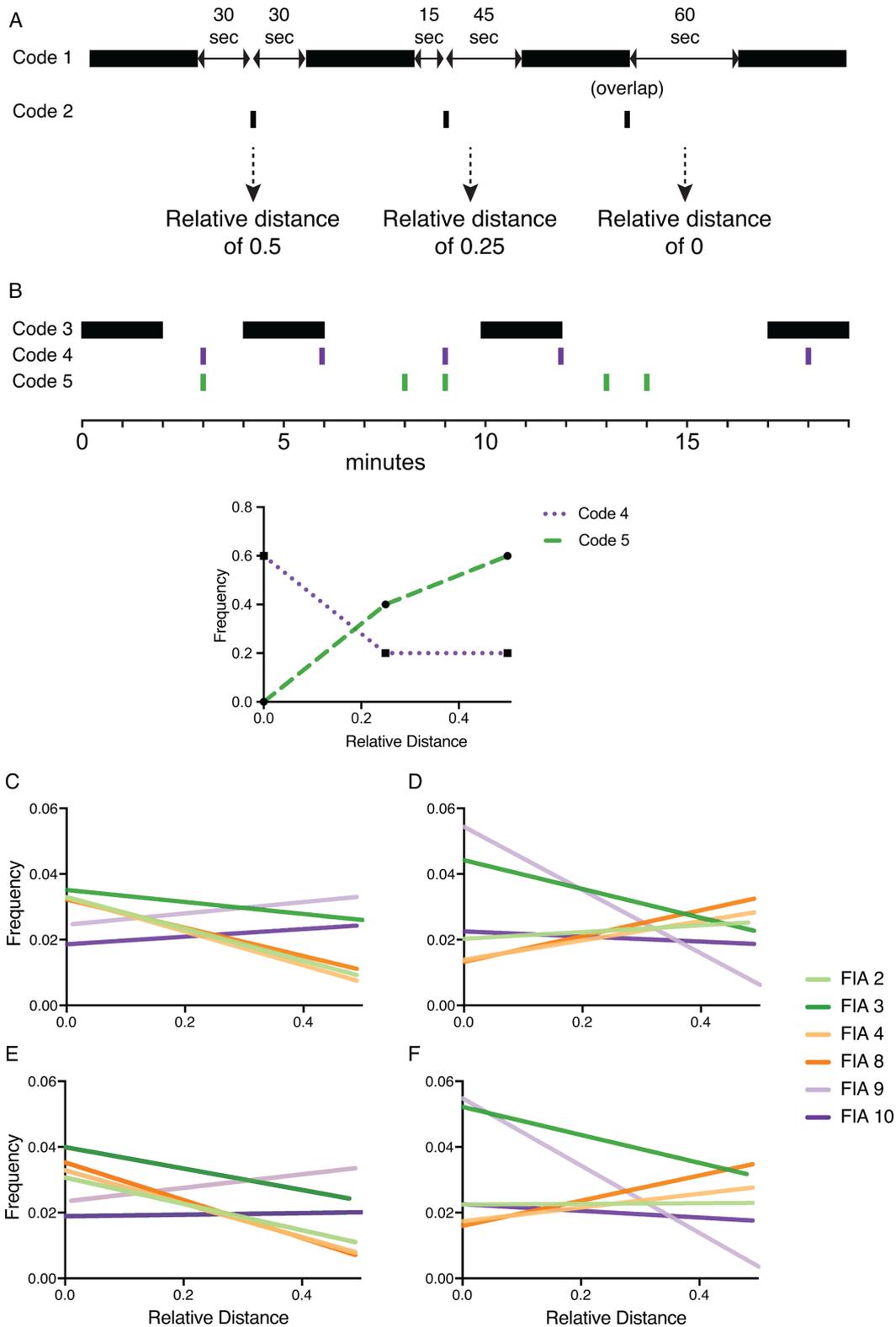


FIGURE 4. Diagrams of relative distance metric and linear regression analysis of the frequency of relative distance values for six FIA codes relative to individual COPUS codes. (A) Three examples of individual relative distance measures. (B) Timeline of three codes, with an accompanying frequency plot for relative distance values for code 4 or code 5 relative to code 3. (C) Plot for COPUS code PQ (Instructor poses question). (D) Plot for COPUS code SQ (Student asks question). (E) Plot for COPUS code SAnQ (Student answers question). (F) Plot for COPUS code IAnQ (Instructor answers question). Abbreviations: FIA 2, Instructor praises or encourages student; FIA 3, Instructor accepts or uses ideas of students; FIA 4, Instructor asks nonrhetorical questions; FIA 8, Student talk in response to a teacher’s question or instructions; FIA 9, Student talk initiated by the student; FIA 10, Silence.

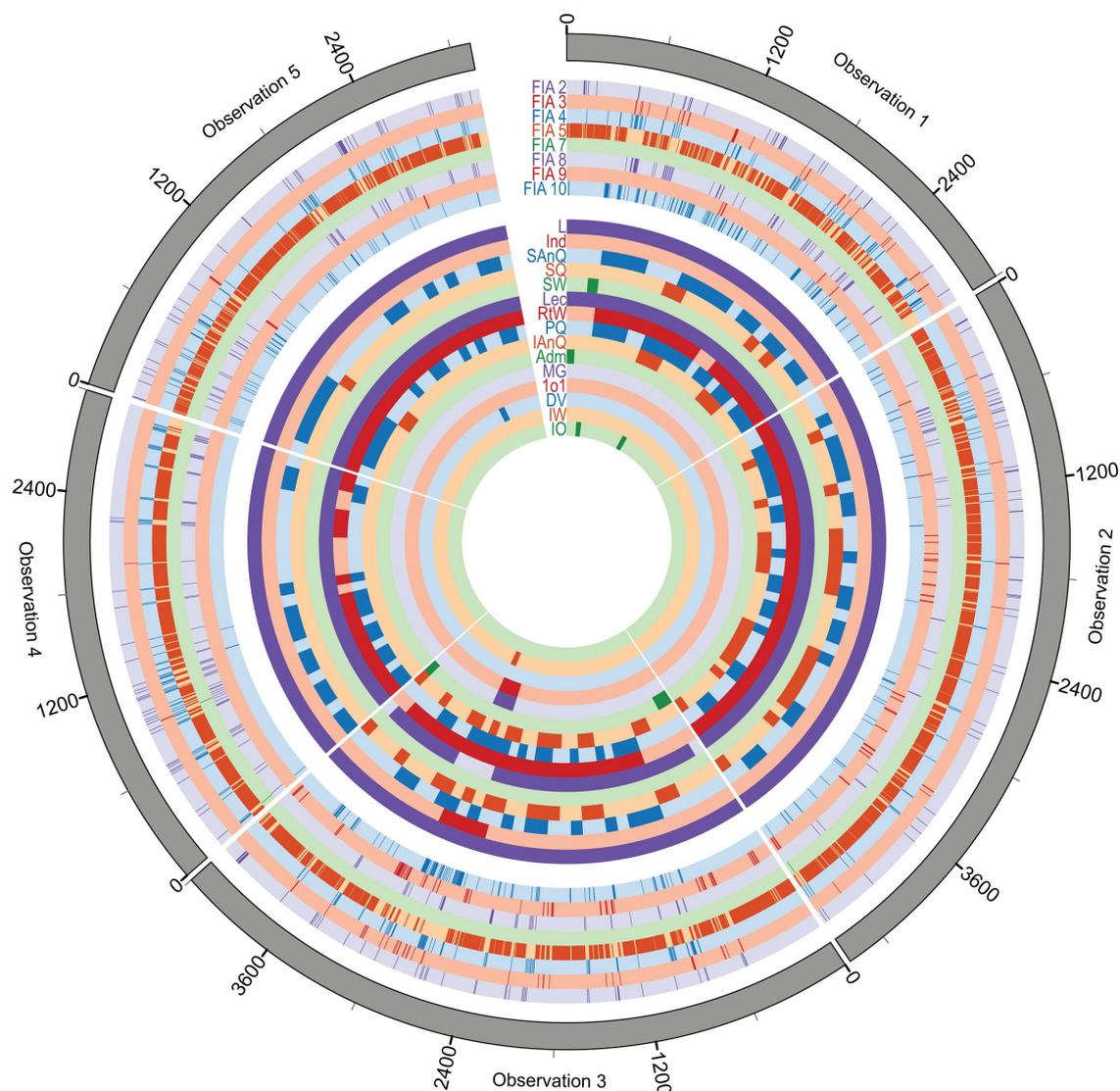


FIGURE 5. Circos plot illustrating all observed COPUS (inner region) and FIA (outer region) codes from observations 1–5 (Table 2). Dark colors within a circular track indicate the presence of a code, while lighter shades of the same color indicate the absence of the same code. Numbers on the outside of the circle indicate the progression of time in seconds for each of the five observations. Abbreviations: FIA 2, Instructor praises or encourages student; FIA 3, Instructor accepts or uses ideas of students; FIA 4, Instructor asks nonrhetorical questions; FIA 5, Instructor lectures; FIA 7, Instructor criticizes student; FIA 8, Student talk in response to a teacher’s question or instructions; FIA 9, Student talk initiated by the student; FIA 10, Silence; L, Students listening; Ind, Student independent work; SAnQ, Students answer questions; SQ, Students ask questions; SW, Students wait; Lec, Instructor lectures; RtW, Instructor writes on board during instruction; PQ, Instructor poses question; IAnQ, Instructor answers question; Adm, Instructor performs administrative task; MG, Instructor moves through class during group work; 1o1, Instructor talks to a student one on one; DV, Instructor uses a demo/visual; IW, Instructor waits; IO, Instructor engages in an “other” unclassified task.

An example where this might apply to classroom observation data relates to audio recordings of classroom activities, such as those used within the DART protocol (see the Supplemental Material). The volume of a recording can be determined for various spans within the class period. The fourth column would allow each span to be annotated with the appropriate volume value, while remaining within the same timing framework used by bed files not containing the additional column of information.

Analysis Methods and Tools

Once classroom data have been converted to a genomic file format, a variety of tools are available to extract meaning from the raw data. A prime example is the Bedtools suite of software tools (Quinlan and Hall, 2010). Bedtools enables a broad spectrum of analyses upon and between bed files. Some components of the suite have little conceivable use within a CAG context—perhaps unsurprisingly, seeing as how the suite was originally designed for a different purpose. However, many of

the tools are broadly useful even outside genomics. For example, the “bedtools closest” command would allow a researcher to see what codes most closely surround another code of interest as well as the distance separating the codes, while the “bedtools jaccard” command would allow a researcher to calculate the Jaccard statistic for overall similarity between different codes. Through the command line, bedtools-based analyses can be automated and applied to large data sets with ease. There are also options for implementing Bedtools and other bioinformatics packages using a graphical user interface (GUI) as opposed to the command line—Galaxy is an example of a Web-based, GUI-oriented platform for conducting bioinformatics analyses (Afgan *et al.*, 2018). In a broad sense, the creative repurposing of bioinformatics tools and workflows can allow education researchers to avoid reinventing the wheel.

Visualizing Classroom Observation Data in Raw or Analyzed Form

Encoding classroom data in bed file format enables a researcher to visualize codes using the same tools used by biologists to visualize genomics data streams, for the purposes of both initial exploration and final display. A variety of genome browsers are available, such as the Integrated Genome Viewer (Robinson *et al.*, 2011) and the Integrated Genome Browser (Freese *et al.*, 2016), that allow for customized visualization and exploration of files in standard genomic formats; these modes of representation served as the inspiration for Figure 1. Genome browsers excel for the interactive viewing of data and allow for general data exploration. In contrast, command-line tools such as Circos (Krzyszowski *et al.*, 2009) or ends analysis scripts (Picard and Gehring, 2017) are excellent for the static presentation of either raw or analyzed data in figure-ready form, as well as for the construction of multiple related plots through automation. In addition, some observation protocols, such as OPAL (Frey *et al.*, 2016), RPDOT (Olmstead and Turpen, 2016), and RIOT (West *et al.*, 2013), are already associated with visualization schemes that share characteristics with the genomic visualization tools described earlier (Supplemental Material). In any case, modifications might sometimes be necessary to enable the visualization of complementary and/or unanticipated data streams.

CONCLUSION

In this paper, we argue that classroom observation data, as collected using any of a number of observation protocols, can be analyzed in new ways through pairing with established bioinformatics frameworks and tools.

One criticism that is sometimes levied at big data approaches is that they can seemingly favor eschewing research question-driven or hypothesis-driven studies in favor of large-scale “fishing expeditions.” However, as shown through the earlier example analyses, hypothesis-driven research is fully compatible with the CAG approach. Therefore, we see the ability to conduct exploratory surveys of large, complicated data sets as a feature, rather than a bug, as this can lead to a plethora of new research questions that may not have been tractable within traditional analysis regimes. Indeed, researchers in the data analytics community have argued that “Data by themselves have only limited value. True transformation of educational ecosystems lies in

converting these data into actionable intelligence (meaning insights and knowledge that enable learners and other stakeholders to act)” (Madhavan and Richey, 2016, p. 6).

Multiple members of the science education research community have argued that it is important to expand efforts within the realm of second-generation science education research, and such an expansion should include studies that explore the specific aspects of instructor and student behavior that lead to learning gains within active-learning environments (e.g., Freeman *et al.*, 2014; Dolan, 2015). We propose that CAG analyses can play an important role in such second-generation studies by enabling researchers to pinpoint previously unnoticed or subtle patterns in behavior and activity within the postsecondary biology classroom. Just as genomics approaches have allowed for deeper, more nuanced analyses of genetic influences on human disease than were once possible or even imaginable, CAG approaches have the potential to expand the scope of classroom observation protocol-based studies beyond summary statistics, into the realm of detailed pattern analysis.

An additional benefit of the CAG mindset is the potential to improve communication between biology education researchers and bench biologists. A study within physics suggests that the interactions between discipline-based education researchers and instructors are not always smooth (Henderson and Dancy, 2008). Biology education research efforts might be viewed as more compelling and easier to understand by biologists if they are framed using the language of genetics and genomics. Because much of the biology education community works within traditional biology departments, finding new ways to facilitate meaningful connections between education research and biology research could be of great benefit to researchers on both sides.

One of the most exciting aspects of this approach is the potential for multiple observation protocols to be applied to the same classroom observations, allowing for better triangulation during analysis. This approach admittedly requires a measure of caution. Validity, the idea that an observation protocol or other instrument is capturing what a researcher argues it is capturing, is not an inherent or immutable property of an instrument (Campbell and Nehm, 2013). This means that observation protocols designed for different instructional contexts should be closely examined to determine whether or not codes are measuring the same or related constructs. However, if validity is kept in mind during data analysis, differing instruments may be used in parallel to great effect. If classroom video recordings could be coded by multiple collaborating research groups using different protocols, with the resulting annotations available to other researchers, the community would benefit from both “widening” (making more observations available) and “deepening” (annotating each video with multiple, complementary instruments). The creation of some form of a semi-open data ecosystem within the science education research community could catalyze great progress in this area, as data collected with the intention of addressing one question could also be brought to bear upon unforeseen questions and reduce duplication of data-collection efforts. While this would require close coordination among researchers, institutional review board officials, and other stakeholders, the benefits of having an educational equivalent to the Gene Expression Omnibus could be sizable.

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation under grant no. 1552448. We thank A. Kelly Lane and Maia Popova for feedback on previous drafts and the monitoring editor and two anonymous reviewers for their constructive comments.

REFERENCES

- Achen, R. M., & Lumpkin, A. (2015). Evaluating classroom time through systematic analysis and student feedback. *International Journal for the Scholarship of Teaching and Learning*, 9(2), Article 4. <https://doi.org/10.20429/ijstl.2015.090204>
- Afgan, E., Baker, D., Batut, B., Beek, M., Bouvier, D., Cech, M., ... Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(w1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- American Association for the Advancement of Science. (2012). *Describing and measuring undergraduate STEM teaching practices*. Retrieved April 10, 2018, from <http://cclconference.org/measuring-teaching-practices>
- Amidon, E., & Flanders, N. (1967). *The role of the teacher in the classroom*. Minneapolis, MN: Association for Productive Teaching.
- Ancombe, F. (1973). Graphs in statistical analysis. *American Statistician*, 27(1), 17–21. <https://doi.org/10.2307/2682899>
- Campbell, C. E., & Nehm, R. H. (2013). A critical analysis of assessment quality in genomics and bioinformatics education research. *CBE—Life Sciences Education*, 12(3), 530–541. <https://doi.org/10.1187/cbe.12-06-0073>
- Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558), 1306–1311. <https://doi.org/10.1126/science.1067799>
- Dolan, E. (2015). Biology education research 2.0. *CBE—Life Sciences Education*, 14(4), ed1. <https://doi.org/10.1187/cbe.15-11-0229>
- Ebert-May, D., Derting, T., Hodder, J., Momsen, J., Long, T., & Jardeleza, S. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *BioScience*, 61(7), 550–558. <https://doi.org/10.1525/bio.2011.61.7.9>
- Favorov, A., Mularoni, L., Cope, L., Medvedeva, Y., Mironov, A., Makeev, V., & Wheelan, S. (2012). Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Computational Biology*, 8(5), e1002529. <https://doi.org/10.1371/journal.pcbi.1002529>
- Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, 111(23), 8410–8415. www.pnas.org/cgi/doi/10.1073/pnas.1319030111
- Freese, N., Norris, D., & Loraine, A. (2016). Integrated Genome Browser: Visual analytics platform for genomics. *Bioinformatics*, 32(14), 2089–2095. <https://doi.org/10.1093/bioinformatics/btw069>
- Frey, R., Fisher, B., Solomon, E., Leonard, D., Mutambuki, J., Cohen, C., ... Pondugula, S. (2016). A visual approach to helping instructors integrate, document, and refine active learning. *Journal of College Science Teaching*, 45(5), 20–26. https://doi.org/10.2505/4/jcst16_045_05_20
- Henderson, C., & Dancy, M. (2008). Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American Journal of Physics*, 76(1), 79–91. <https://doi.org/10.1119/1.2800352>
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... Marra, M. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9), 1639–1645. <https://doi.org/10.1101/gr.092759.109>
- Lund, T., Pilarz, M., Velasco, J., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE—Life Sciences Education*, 14(2), ar18. <https://doi.org/10.1187/cbe.14-10-0168>
- Madhavan, K., & Richey, M. (2016). Problems in big data analytics in learning. *Journal of Engineering Education*, 105(1), 6–14. <https://doi.org/10.1002/jee.20113>
- Olmstead, A., & Turpen, C. (2016). Assessing the interactivity and prescriptiveness of faculty professional development workshops: The real-time professional development observation tool. *Physical Review—Physics Education Research*, 12(2), 020136. <https://doi.org/10.1103/PhysRevPhysEducRes.12.020136>
- Picard, C., & Gehring, M. (2017). Proximal methylation features associated with nonrandom changes in gene body methylation. *Genome Biology*, 18(1), 73. <https://doi.org/10.1186/s13059-017-1206-2>
- Quinlan, A., & Hall, I. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Robinson, J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E., Getz, G., & Mesirov, J. (2011). Integrative Genomics Viewer. *Nature Biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Smith, M., Jones, F., Gilbert, S., & Wieman, C. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, 12(4), 618–627. <https://doi.org/10.1187/cbe.13-08-0154>
- Smith, M., Vinson, E., Smith, J., Lewin, J., & Stetzer, M. (2014). A campus-wide study of STEM courses: New perspectives on teaching practices and perceptions. *CBE—Life Sciences Education*, 13(4), 624–635. <https://doi.org/10.1187/cbe.14-06-0108>
- Stains, M., Harshman, J., Barker, M., Chasteen, S., Cole, R., DeChenne-Peters, S., ... Young, A. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470. <https://doi.org/10.1126/science.aap8892>
- West, E., Paul, C., Webb, D., & Potter, W. (2013). Variation of instructor-student interactions in an introductory interactive physics course. *Physical Review Special Topics—Physics Education Research*, 9(1), 010109. <https://doi.org/10.1103/PhysRevSTPER.9.010109>