Published in final edited form as:

J Sci Educ Technol. 2016 February ; 25(1): 91–110. doi:10.1007/s10956-015-9579-z.

Using Interactive Data Visualizations for Exploratory Analysis in Undergraduate Genomics Coursework: Field Study Findings and Guidelines

Barbara Mirel¹, Anuj Kumar², Paige Nong³, Gang Su⁴, and Fan Meng⁴

Barbara Mirel: bmirel@umich.edu

¹School of Education, University of Michigan, Ann Arbor, Michigan

²Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan

³Leondale Germiston, South Africa

⁴Molecular and Behavioral Neuroscience Institute, University of Michigan Medical School, Ann Arbor, Michigan

Abstract

Life scientists increasingly use visual analytics to explore large data sets and generate hypotheses. Undergraduate biology majors should be learning these same methods. Yet visual analytics is one of the most underdeveloped areas of undergraduate biology education. This study sought to determine the feasibility of undergraduate biology majors conducting exploratory analysis using the same interactive data visualizations as practicing scientists. We examined 22 upper level undergraduates in a genomics course as they engaged in a case-based inquiry with an interactive heat map. We qualitatively and quantitatively analyzed students' visual analytic behaviors, reasoning and outcomes to identify student performance patterns, commonly shared efficiencies and task completion. We analyzed students' successes and difficulties in applying knowledge and skills relevant to the visual analytics case and related gaps in knowledge and skill to associated tool designs. Findings show that undergraduate engagement in visual analytics is feasible and could be further strengthened through tool usability improvements. We identify these improvements. We speculate, as well, on instructional considerations that our findings suggested may also enhance visual analytics in case-based modules.

Keywords

Data visualization; Case-based instruction; Critical thinking; Technology; Visual analytics; College biology

Correspondence to: Barbara Mirel, bmirel@umich.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s10956-015-9579-z) contains supplementary material, which is available to authorized users.

Introduction

Life science researchers increasingly engage in visual analytics to explore high volume, multidimensional-omics data for novel insights. In visual analytics, researchers “apply[] analytical reasoning to interactive visual interfaces” to find previously unknown relationships in perceived patterns and other data (Keim et al. 2010, 7). For example, with tools like interactive heat maps, life scientists can dynamically examine transcriptional profile data; statistically cluster and functionally group differentially expressed genes; and filter, subset and compare data to uncover biologically meaningful relationships (Gehlenborg et al. 2010). The goal is to generate a hypothesis about molecular-level mechanisms of phenotype-level traits or disease. In college, life science majors should be engaging in the same practices as scientists (Anderson et al. 2012). But visual analytics for the exploration of large-scale data is one of the most underdeveloped areas of teaching and learning in undergraduate biology today (Iskopehi 2014; Lee and Tsai 2013).

A joint NSF-NIH report emphasizes that interactive visualizations play a significant role in scientific reasoning, noting: “Understanding and, ultimately, knowledge cannot be delivered directly from computation. Visualization is the tool through which computation presents a face to an end user and, by so doing, it allows the user to derive knowledge from data... [and] extend human capabilities” (Johnson et al. 2006, 5). Many science classrooms engage students with scientific and structural visualizations. Far fewer give them experience with abstract data visualizations for exploratory analysis. Yet data visualizations promote the same kind of higher-order reasoning envisioned by the new twenty-first century biology curricula (Johnson et al. 2006). Such higher-order reasoning includes recognizing and drawing meaning from patterns, breaking problems into parts, making choices based on reasoned arguments, determining the quality or importance of information, and weighing the relative value of evidences (Lemons and Lemons 2013). Visual analytics can help students develop and enhance such reasoning skills. Yet it must be feasible for students to use such tools. That is, tools have to have relevant features and functions for the complex analysis assigned in class, and they have to be usable for undergraduates. Thus, a first step in introducing visual analytics to undergraduate biology majors is to establish whether it is feasible for them to use data visualizations for complex tasks and problems and, correspondingly, to identify usability features that need to be in place for conducting these tasks and problems. Our study seeks to establish feasibility, to identify student successes and difficulties with visual analytics and to uncover usability issues.

Empirically grounded usability standards have long played a role in the design of graphic user interfaces (Nielsen 1993). By comparison, standards for usable data visualizations are still in their formative stages, and relatively few have been established empirically (Scholtz 2011). Of the usability standards that exist, those focused on cognitively facilitating higher-order reasoning for complex analysis are particularly relevant to studying visual analytics in an undergraduate biology class. For example, current usability designs emphasize that interactive visualizations with side-by-side displays facilitate complex (multi-factor) reasoning about relationships and multiple perspectives better than displays arranged in separate tabs (Gehlenborg et al. 2012). Usability studies of interactive visualizations also show that the content more than the volume of displays affects analysts’ ease of meaning-

making. Dense screens, if they are filled with relevant data targeted to analysts' intentions, facilitate meaning-making better than low-volume data displays that contain extraneous/irrelevant content (Lavie 2000). Irrelevant data distract analysts' attention and obstruct sensemaking. Additionally, usability standards stress that visualizations need to perceptually cue potential data of interest in order to afford seamless flows of thought and interpretation in complex reasoning. Analysts' reasoning also benefits from visualization designs that couple interactive graphics with labels and details about the data (Saraiya et al. 2005a). Finally, usability studies show that analysis is improved when visualizations are integrated with built-in statistical calculations (Peer and Shneiderman 2008).

These usability findings pertain to data visualizations used by professional scientists. At present, little empirical research exists to demonstrate whether it is feasible for undergraduates to use the same data visualization applications as professionals do and whether these tools are usable for students. Engagement with the same tools is important because by using the same tools students can participate in the same exploratory analysis practices as professional scientists—albeit with differing levels of domain and methodological expertise. Fortunately, a handful of studies have begun to evaluate biology students' uses of the same data visualization tools that professional scientists use. One such study in a middle school biology classroom reveals promising results, showing that the data visualizations improved student performance and learning. The researchers of this study tied students' successes to usable visualization designs—to designs that “support[ed]... different [nonlinguistic] representations, ...provide[d] ...flexibility,...and stimulate[d] greater mental effort and more refined ideas” (Oviatt and Cohen 2010, 516; Peters and Songer 2013). The researchers argued that students' visual analytic achievements likely would have been even greater if the visualizations were additionally designed with the following usability features: (a) visual cues to draw students' attention to task-relevant relationships in the problem space and (b) opportunities to take different perspectives on the same data to facilitate deeper understanding (Peters and Songer 2013; Levy 2013; Gehlenborg et al. 2010).

For undergraduate biology, studies on the uses and usability of data visualizations in classroom curricula are more anecdotal (Shubert et al. 2009; Ngyuen et al. 2012). Anecdotally, studies highlight a connection between students' task-driven visual analytic performances and such usability features as: (a) giving students quick access to external sources of information; (b) providing pre-defined views for taking multiple perspectives of microarray data; and (c) providing pre-defined calculations so that students can engage in the same statistical reasoning and data transformations as scientists do (Shubert et al. 2009). In both the middle school and undergraduate instances, findings echo the usability standards mentioned earlier as means to facilitate higher-order analysis. A closer look at knowledge and skills relevant to the specific analysis in our field study is presented in the next section.

The need for more insight into issues of feasibility and usability of data visualizations for undergraduates majoring in biology comes at an opportune time. Because data visualization software for biomedical exploratory analysis is presently in its early stages, insights into students' readiness for exploratory visualizations can both help define design requirements for these emerging applications and help guide classroom instruction in visual analytics. To examine feasibility and usability in relation to the skills and knowledge involved in a

course-based visual analytics module, we conducted a two-and-a-half-week field study of 22 biology majors in an upper-level genomics class as they engaged in a pilot module. Working in pairs, students interacted with an exploratory (professional-level) tool for generating dynamic heat maps called CoolMap, an open-source, NIH-funded application developed by two of the authors (GS and FM) (Su 2013). Static heat maps are conventional graphics used in transcriptional profiling studies. They give color-coded representations of transcript levels for a given set of genes in various samples or in response to treatments to help life scientists facilitate perceive and cognitively process expression trends across a transcriptome. Static heat maps, however, cannot capture in a single screen or sheet of paper high-volume data sets, for example those with thousands of genes (rows) and scores of conditions or samples. Students in the field study worked with such a large data set and used CoolMap to interactively explore and find meaningful biological relationships. Their goal was to find and to functionally and statistically explain treatment-expression effects that could provide biological insight into the genetic basis of relevant signaling pathways or cell processes. We analyzed student performance to assess the feasibility of this tool for undergraduate course application. To do so we determined the relative distribution of time students spent performing exploratory tasks and time spent interacting with the basic program operations that enabled navigation, setup and analysis. Within the time that each pair spent in program operations, we determined the percentage of time a pair spent on each program operation. We also determined the extent to which students achieved task completion. Additionally, we examined student successes and shortcomings in visual analytic thinking processes. We defined knowledge and skills relevant to the transcript profiling module by first abstracting from the literature on visual analytics the reasoning and sensemaking that data visualizations demonstrably afford. Then, we mapped these processes to constituent knowledge and skills specified in the literature on the New Biology curricula. We further informed these knowledge and skills by drawing on research literatures on cognition for inquiry and higher-order thinking, as well. Focusing on knowledge and skills germane to the transcript profiling module, we analyzed field study data to characterize the extent to which students applied relevant knowledge and skills as specified in five categories found in the New Biology research literature. These categories include: Prior domain knowledge, nature of science knowledge, visual literacy and visual thinking, statistical reasoning and higher-order thinking. For insight into usability designs that may have improved students' visual analytics, we assessed students' demonstrated shortcomings, gaps and errors in relation to tool design.

Field study results show that it is feasible for undergraduate life science majors to explore large data sets with a professional-level dynamic heat map tool for discovery-driven inquiries. In our case, certain usability improvements are needed to assure this feasibility, and we present them. Additionally, field study results indirectly suggest possible instructional improvements for strengthening students' skills in exploratory analysis.

Background Knowledge and Skills for Visual Analytics

Because few if any studies directly identify skills and knowledge required for undergraduate-level visual analytic explorations, we abstracted implicit criteria for biology majors' success in visual analytics from the variety of research sources mentioned

previously. Our synthesis shows that required knowledge and skill are distributed across five areas, as follows.

- *Prior domain knowledge.* The New Biology curricular standards identify many areas of domain knowledge that were relevant to the exploratory inquiry in our field study case. They include understanding genomics, sequence variation, regulatory processes and pathways. Specifically, students needed to know “how mutations in the DNA sequence of a gene ... result in phenotypic change; ...the relationship of genes to proteins and phenotypes;... and genes’ influences [on]... functions and behaviors of a living organism” (Wefer and Sheppard 2008, 158–159). Relevant knowledge also involved understanding “interactions among molecules ... and functional activities that emerge as causal mechanisms within or between cells;” effects of mutations or variations, and “gene regulation (turning genes on and off at the right time)” (Trujillo et al. 2012, 101; Bowling et al. 2008). Additionally, relevant to our field study, the New Biology standards stress the need for students to be able to develop and fill out mental models of multi-faceted molecular events and “behaviors in molecular systems including, paradoxical behaviors” (AAAS 2011; NRC 2009). Finally, another crucial aspect of domain knowledge was familiarity with the data being investigated. (Saraiya et al. 2005b).
- *Nature of science (NOS) knowledge.* NOS knowledge refers to various aspects of “doing science.” Relevant curricular standards for NOS knowledge for our field study include knowing when and how to validate findings; thinking reflectively about encountered and self-produced knowledge; and differentiating between opinion and evidence (Sadler and McKinney 2010, 43; Sadler et al. 2000; NRC 2010). NOS knowledge also includes realizing that insights in discovery do not come quickly or easily; they take patience and perseverance.
- *Visual literacy and interactive visual thinking.* For the field study, students needed to be able to read information off the graphic and understand it. Additionally, as advocated by the New Biology curricular standards, students needed to know and notice what counted as a pattern in the graphics, to assign patterns to conceptual categories, and to interpret biological meaning from displayed relationships. These visual thinking skills depend on students being able to strategically manipulate displays for sensemaking, i.e., knowing when and how to adjust layouts, arrange and re-arrange data, and move between levels of detail (Oviatt and Cohen 2010; Liu et al. 2008).
- *Statistical reasoning.* Many visualized relationships have statistical underpinnings, and students needed to be quantitatively literate as specified by the New Biology curricular standards, e.g., understanding central tendencies and covariance (Smolinski 2010). Additional standards relevant to the statistical reasoning students needed for the visual analytics in the field study include being able to draw statistical inferences from data sets and to deduce rules that govern patterns (NRC 2003; Colon-Berlingeri and Burrowes 2011). They also include drawing inferences from statistical relationships to address a biological question (Colon-Berlingeri and Burrowes 2011).

- *Higher-order thinking.* For the visual analytics in the field study module, relevant higher-order thinking from the New Biology curricula standards include students being able to take into account and understand interactions among several factors, conditions and events; and taking different visual and conceptual perspectives of the same data (Songer et al. 2009, 614; Lengler 2006). They include, as well, being able to tackle cognitive dissonance, which often arises due to seemingly paradoxical biological relationships (Spiro et al. 1992).

Methodology

To examine undergraduates' visual analytic behaviors, reasoning and outcomes in an upper-level genomics course, we implemented an interactive heat map designed for professional exploratory analysis. We audio- and video-captured students' performances. We analyzed performance patterns and strengths and gaps in the knowledge and skills students applied. We examined tool-related (usability) obstacles linked to errors and to gaps in aspects of knowledge and skill related to exploratory analysis with data visualizations to identify needed usability improvements. We also inferred instructional considerations for enhancing visual analytics in college.

Software

CoolMap is a dynamic heat map coupled with hierarchical ontologies that users create to characterize a data set. By applying ontologies, users can display data in aggregated categories, e.g., by Gene Ontology terms or disease classifications. Through ontologies, users can group big and complex -omics data into much smaller and clearer views, thereby condensing unwieldy data sets into biologically meaningful categories. For displays of biologically meaningful ontology groupings, CoolMap color codes the cells by select summary values, e.g., mean, median or standard deviation. By looking at a variety of summary values, users get a sense of how individual values (e.g., fold changes) are distributed in an aggregate. After users identify high-level patterns in aggregated displays, they can drill down to further explore individual-level data.

Users can dynamically manipulate rows and columns to explore individual-level or aggregated data. Users also can statistically cluster rows, columns or both. Additionally, CoolMap includes numerous tabbed panes surrounding the main heat map window (see Fig. 1).

Functions and features in these tabs give users interactive means for searching, filtering, changing color coding, altering color coding parameters, adding data and accessing detail on the data. Given these displays and interactions, users can apply their knowledge to identify novel and potentially biologically meaningful patterns and insights for follow-up statistical analysis and experimentation. CoolMap was developed in Java using a modular design and custom-developed high-performance and extensible rendering engine—e.g., it redraws a 6000×6000 cell heat map in less than a second.

Student Profile

Of the 22 students in the study, one was a graduate student, 19 were seniors (graduating at the end of the month) and two were juniors. Seven were female; 15 were male. Three were honors students (all males). All were biology majors.

Module Development and Focus

The genomics professor (AK) developed the module over several iterations, while continuously consulting with the CoolMap developers (GS and FM) and user experience/instructional design specialist (BM). The content and purpose of the module fit squarely into the course curriculum. The curriculum investigated methods for sequencing genomes and for analyzing gene and protein functions on a large scale across a diverse spectrum of eukaryotes.

The module ran at the end of the semester, enabling students to build on prior course learning. It asked students to find and describe strong gene expression changes in response to various perturbations (gene mutations and chemical treatments). It also asked students to explain possible biological implications of treatment-induced expression changes. The module was grounded in the same active learning methods that the professor used throughout this course and that he uses in other courses, as well (Kumar 2005). To prepare for hands-on visual inquiry, the students read and discussed a primary source article—Hughes et al. (2000). It served as the basis for the inquiry case and provided the data set that students explored with CoolMap. The data set identified transcript levels for each of approximately 6000 genes in *Saccharomyces cerevisiae* in response to 300 perturbations.

To guide the case-based exploration, students were given a user manual for CoolMap and a written instructional guide (see, respectively Supplemental Material 1 and 2). The written instructions mixed structured directions (prompted tasks) with questions that encouraged unspecified (unprompted) tasks relevant to the objectives. The written instructions moved students through three tasks, which built on each other successively. The tasks were structured by a learning progression adapted from the research literature (Songer et al. 2009). The progression advanced students through inquiries that moved from simple to more complex core disciplinary concepts, crosscutting concepts and reasoning practices. The written instructions also asked students to answer three questions for each task and to record answers on a separate worksheet. Two of the questions asked for read-offs, namely for the names of treatment-transcript relationships with strong expression changes. The third question asked students to biologically explain one of the strong relationships.

Descriptions of each task, its subtasks and intellectual demands are summarized in Tables 1, 2, and 3.

Task 1: erg Genes—*Find and describe salient patterns and identify strong treatment-gene effects specifically in ergosterol biosynthesis (erg) genes* (see Table 1). Students first were to visually locate and name visual patterns, e.g., serial cells in rows or columns with similarly high or low fold change colors. Such pattern recognition is fundamental to visual analytics. They then were to select and display a pre-defined functional aggregate of genes

—those belonging to the ergosterol (erg) biosynthesis pathway—thereby honing in on meaningfully classified items for relational reasoning. Through the perceptual processing and visual literacy afforded by data visualizations, students were to find treatments that had strong effects on the erg aggregate as indicated by the color coding. Having found strong summary fold changes based on the color coding for mean values, students then were to drill into individual transcripts within the erg aggregate to find genes with the strongest expression changes in response to a treatment. Drilling down reinforced statistical understanding by visually revealing (through color) distributions of values represented in statistical means. From the individual-level fold changes students next were to focus on an apparent strong expression change and try to explain it speculatively, taking into account its ergosterol biosynthesis function. This speculation was students' initial experience in the module constructing an event-based biological explanation. This task consisted of the following subtasks and intellectual demands—some prompted, others implicit in the task but not prompted.

Task 2: CDC Genes—*Within an up- or down-regulated functional subgroup of cell division cycle (CDC) genes, locate, name, and explain the strongest expression change in an individual gene in response to a treatment* (see Table 2). Students added a new functional grouping to the heat map—this time an aggregate of cell division cycle (CDC) genes. This aggregate, in turn, had two sub-groups: up-regulated genes and down-regulated genes. Half the students focused on the up-regulated CDC genes and the other half on the down-regulated CDC genes. For expression changes—minimum, maximum, median, first quartile, third quartile, standard deviation, and variance—students recorded the heatmap cell color, representing the aggregate value. Ideally, by viewing these different statistics, students would infer how member values in the aggregate might be distributed. After finding a treatment tied to a strong CDC expression change at a summary level, students drilled into the aggregate-treatment relationship. As in the erg task, they found the individual transcript with the strongest expression change and explained it biologically. Now, however, they were prompted explicitly to examine outside information for their explanation. Some subtasks repeated those from Task 1. Distinct subtasks were as follows:

Task 3: Hierarchical clusters of genes—*Within a hierarchically clustered view of the dataset, locate and describe a cluster with strong fold changes that has one unknown (“y”) gene and project next explanatory steps* (see Table 3). Students did not complete this task, which—in the structure of our adapted learning progression—was intended as a final progression toward higher-order thinking. The time that the syllabus allotted to the module ended before students could begin this task. This task involved viewing the data in unbiased clusters, thereby engaging students in a different perspective from the functionally (ontologically) grouped heat map views in the prior tasks. In the third task, students were to find a cluster that included at least one gene of unknown function. They then were to propose the next investigative steps for functionally explaining individual gene–treatment relationships in the cluster and possible descriptions of the functionally uncharacterized in light of these other relationships in the cluster. They also were to compare clustering and functional grouping as methods for finding and biologically understanding strong relationships. Subtasks and intellectual demands were as follows:

Students spent five classes over two and a half weeks on the module. The class met twice a week for one and a half hours each class. Table 4 shows how students actually spent their time, with no time left for Task 3.

Research Questions and Methods

The field study examined four related research questions:

1. *Feasibility.* To what extent is it feasible for undergraduate life science majors to interact efficiently with a professional-level visual analytics tool to complete inquiry tasks, including open-ended exploration?
2. *Successes.* What behaviors and thinking associated with relevant knowledge and skill for visual analytics do students demonstrate?
3. *Difficulties.* What omissions or difficulties in relevant knowledge and skill do students demonstrate that impede efficient, complete and comprehensive performance?
4. *Usability.* In what ways, if any, are omissions and difficulties tied to tool usability?

Data Collection and Processing

Students worked in pairs on a laptop onto which CoolMap and Camtasia Studio 8.0 were loaded. Camtasia collected time-stamped audio and video recordings of students' work. Unexpected technical difficulties with Camtasia kept us from finalizing some tapes. We were able to collect quality tapes on ten of the 11 pairs for the erg gene task and for eight of 11 pairs for the CDC gene task. We collected answers to worksheet questions for all 11 pairs at the end of the module for further data on students' thinking. Additionally, at the end of the module, all 22 students individually filled out a 25-question anonymous satisfaction survey (see Supplemental Material 3).

We did not collect data on students' prior knowledge, attitudes toward science, prior visual literacy, statistical reasoning or high-order thinking skills. This decision means we were not able to compare pre- and post-abilities and behaviors. This decision was largely practiced. The amount of time allotted to the module in the syllabus was tight, and we wanted to give students as much time as possible to conduct the visual analytics.

Three qualitative researchers (BM, PN, and a consultant) watched the tapes several times alone and jointly. The audio portions of the observations were transcribed. Using holistic analysis methods, the researchers iteratively and jointly derived themes, patterns in student behaviors, and exceptions (Creswell and Clark 2010). They collaboratively developed and agreed on a 34-item coding scheme representing the various behaviors observed in the tapes (see Supplemental Material 4). Fourteen codes dealt with students' program operation behaviors in order to derive durations and frequencies for these behaviors. We expected duration and frequency of such behaviors to address the first research question about feasibility. They would indicate students' patterns and the extent of shared efficiencies—in interacting with the program to achieve analytical objectives. The assumption was that if students demonstrated common measures in performing tasks it would indicate feasibility.

Eleven of these 14 codes pertained directly to instances in which students operated features or functions of the program (e.g., the code “Focus” was applied to instances in which students hovered the mouse over a heat map cell to show information about it). Two other coded behaviors for program operations pertained, respectively, to making and resolving errors. Error-related behaviors also would speak about efficiency in moving through the tasks and correspondingly the feasibility of using interactive visualizations. The fourteenth code for program operation behaviors pertained to commenting on tool problems or needed improvements, another mark of efficiency, presuming that expressed dissatisfaction indicates situations in which the tool impedes workflow progress. Overall, long durations in program operations that should have been fluid (e.g., Zoom or Interact with the Data Matrix), errors and problems also gave insight into the fourth research question—how behaviors may have been tied to the usability of the tool.

Eighteen codes of the 34 codes represented students’ demonstrated thinking behaviors, and they addressed the second and third research questions—students’ successes, difficulties and omissions in engaging in visual analytics for the inquiry tasks and subtasks embedded in the assignment. These 18 codes covered (verbal) behaviors in which students clarified, verified, compared, judged relationships, discussed criteria for judgments, monitored knowledge, and negotiated meaning (e.g., “Discuss color shades”). These codes enabled us to analyze durations and frequencies of different types and levels of reasoning within and across pairs. They shed light on students’ visual analytic successes and difficulties. Of the remaining two codes in the 34 code scheme, one tagged time in silence without activity; and the other pertained to students’ references to instructions (“Active with tactics/instruction”).

The three researchers coded practice tapes first for reliability and achieved a Kappa score for reliability in duration coding of 0.91 and for reliability in frequency coding of 0.71.

Data Analysis

The researchers analyzed data with mixed qualitative and quantitative methods and triangulated outcomes from observed and coded sessions, written worksheets, and satisfaction survey response (Creswell and Clark 2010). Qualitatively, they abstracted themes and representative passages from the recorded observations to guide the coding of behaviors. Quantitative measures were calculated from coded behaviors for durations, proportionate durations and frequencies. The researchers also textually analyzed written explanations in worksheets. The textual analysis differentiated explanations based on the number of factors students considered and the extent to which they related factors to the event and conditions of expression change (Crow et al. 2008). The more factors and relationships students’ included, the more complex and higher order the explanation was.

Behaviors coded “Active with tactics/instruction” (reference to instructions) were omitted from the data analysis. For all pairs, the durations and frequencies of this behavior were far higher than those of other codes, and they skewed results. Omitting these data was fitting since students’ references to instructions were not directly related to skills and knowledge associated with visual analytic proficiency. Time spent on this behavior, however, did indicate a tight coupling between instruction and visual analytic performance. A systematic analysis of instruction was outside the scope of our research. But given the prevalence of

students' references to instructions, we inferred from results some instructional considerations that might enhance students' visual analytic performances.

For the first research question—"To what extent is it feasible for undergraduates to interact efficiently with professional level visual analytic tools for inquiry?"—data analysis involved calculating total and mean durations of various program operations as proxies for "efficiencies." Task completion rates were calculated by assessing the percent of students who completed each of the six worksheet questions, three for each task. The benchmark was set at 80 %—the equivalent to the module achieving a "B" average in completion. At 80 per cent task/question completion, it could be inferred that students were able to work the program well enough to accomplish the problem-solving goal for the tasks. Recognizing that efficiencies and task completion rates are likely to be sensitive to error experiences, an analysis of mean times in error and error resolution was conducted and relevant satisfaction survey responses were tallied.

For the second and third research questions—respectively, "What behaviors associated with relevant knowledge and skill did students demonstrate and what difficulties or omissions did they demonstrate?"—data analysis focused on assessing the extent to which students applied behaviors relevant to the knowledge and skills associated with visual analytics derived from the research literature. Students' demonstrated thinking behaviors were especially relevant in shedding light on pairs' applied domain knowledge, NOS knowledge, visual literacy and thinking, statistical reasoning, and higher-order thinking. Results from the textual analysis of written explanations further characterized students' higher-order thinking.

For the fourth research question—"In what ways if any are omissions and difficulties tied to tool usability?"—students' previously mentioned behaviors (especially difficulties/inefficiencies) and the satisfaction survey responses were related to associated tool designs and uses. Anonymous satisfaction responses could not be cross-referenced with specific pairs' performances but provided insights from a different perspective. Results revealed tool problems, necessary improvements and potential instructional enhancements.

Results

Question 1: Feasibility. To What Extent was it Feasible for Undergraduate Life Science Majors to Interact Efficiently with a Professional-Level Visual Analytics Tool for Inquiry Tasks That Included Open-Ended Exploration?

For the first and second tasks, the erg and CDC gene tasks, respectively, students' completion rates were high. Students completed all the worksheet questions at or above the 80 % benchmark except for one—the question asking them to explain a strong treatment-expression change for an erg gene. Only seven of 11 pairs (64 %) completed this question. None of the pairs began or completed the third task. Efficiency clearly did not extend to being able to get through the inquiry demands of all three tasks within the time, instruction and tool bounding the assignment. Within the scope of the first two tasks, however, measures reveal insights into this research question about feasibility.

A majority of students had similar mean durations for 12 program operation behaviors. They include: Zooming, interacting with the aggregator to change summary statistics, searching by name for genes or treatments, focusing on hover-over information on heat map cells, panning, interacting with the Data Matrix to read expression changes in table form, arranging the workspace, using the Radar to see an overview and details of the heat map, navigating to strong CDC expression changes, interacting with the Ontology table to find erg and CDC aggregates and read them into the heat map, arranging the heat map, and resolving errors.

Given that mean durations across students often had ranges, we analyzed patterns in how student pairs apportioned their total time on all program operations by each individual program operation. As Fig. 2 shows the majority of students (at least 62 % for each behavior) apportioned their time similarly for the 12 program operations. These commonalities signal that most students moved through program operations at a shared efficiency level.

One striking exception was time spent in error situations (“Making errors”). All pairs experienced at least one error except for Pair 11. As Fig. 2 shows, Pairs 2, 5, 6 and 8 spent longest in errors, distinctively spending 5 min or more of their total inquiry time in error. Pair 5 spent a long time in just one error due to misreading the instructions and pursuing an errant path for 7 min before realizing it.

Some pairs who spent long times in error also had excessively long durations in other program operations. Pair 2, for example, spent 8 min total in ten types of errors and also had an excessively long duration in arranging the heat map and labels. Errors engendered negative feelings. The four student pairs who spent the longest times in error plus one other pair (Pair 10, who spent the longest time resolving error) made up 58 % of all the negative comments (the behavior coded as commenting on tool problems or needed improvements). Specific errors are detailed in the Usability section.

High error times seemed to be linked to students’ task completion and to the comprehensiveness of their recorded read-offs in the worksheets. Likely, time spent in errors took away from the time students had for task performance. The four pairs who spent the most time in error either did not complete all the worksheet questions or uncovered and entered the lowest number of strong relationships for the erg task.

Questions 2 and 3. Successes and Difficulties. What Behaviors and Thinking Associated with Visual Analytic Proficiency Did Students Demonstrate? What Omissions or Difficulties Impeded Visual Analytic Performances?

Results show the applied knowledge and skills in domain knowledge, nature of science, visual thinking, statistical reasoning and higher-order thinking that entered into students’ visual analytic performances. Verbal conversations between pairs and some coding results for the thinking/reasoning behavior codes give evidence of students’ knowledge and skill. Notably, mean durations of all thinking per pair was considerably less than for program operations. For every 5 min spent operating the program, students spent 3 min thinking.

Domain Knowledge—Students demonstrated several successes in calling up and applying domain knowledge. The case implicitly required students to have and apply domain knowledge about: (1) The fundamentals of functional molecular biology; (2) molecular responses to perturbations; and (3) the data set and species. Except for data set and species, these undergraduates demonstrated a strong grasp of this knowledge. For example, as they worked, they explained processes by which genes turned on and off, effects of mutations, and biological functions of various genes and cellular components. They discussed treatment-transcript relationships in terms of functions and subcellular locations (e.g., “It’s not a regulator;” “It’s a kinase;” “The ergosterol genes are involved in the cell wall.”). Without instructional prompts, the pairs also talked and wrote about treatment effects on gene expressions in terms of inhibition, activation, transcription, translation and pathways. Pairs 9 and 5 for example, discussed why a mutant (treatment) might have the displayed effect on expression:

Pair 9

Student 1: It [erg11] looks like a protein. It catalyzes demethylation in the ergosterol biosynthesis pathway. Student 2: Okay, so if it’s part of that pathway it makes sense that mutating it would really change things.

Pair 5

Student 1: Maybe if cdc23 is more present it means that the drug is activating it, right?

Student 2: I guess so. Or it could be activating something else in its pathway.

Some of the written worksheet explanations also reflected knowledge of functional molecular biology and the processes of regulatory disturbances. But fewer students demonstrated such knowledge in text than in talk.

In contrast with these successes, students were unfamiliar with the data set or with *Saccharomyces cerevisiae* and did not gain familiarity from the brief exposure to the source article. During one of the whole class discussions within the module, several students noted that they would have been able to interpret and explain the relationships in the case had they known more about the species and the data. Instead, students’ interpretations and explanations were fairly generic, e.g., saying that increased expression likely occurred from an activator of expression being deleted or repressed. Had students from the start been better versed in relevant yeast-specific aspects of genomics, they may have been able to deliberately look for and more concretely discuss such issues as: Effects that disruptions in the beginning of the ergosterol biosynthesis pathway have on later pathway processes; implications of simultaneously over-expressed ergosterol genes on sterol composition and the effects of altered composition; and outcomes of disrupted expression in cell division cycle genes on interrelated cell cycle events, cell size and growth. Students also may have been able to more precisely associate mutant-based cell division and ergosterol biosynthesis processes with the inhibited or stimulated synthesis of cell wall components and draw implication for drug targets.

Nature of Science—The area of NOS knowledge in which students were most successful was self-reflection. Across the board, students reflected on their ongoing analyses and evolving knowledge. They engaged in this behavior mostly without direct prompting. For example, as one pair of students began inquiring into biological explanations of a high-expression fold change associated with a particular treatment, they remarked to each other, as follows:

Student 1: They're asking why it makes sense biologically. And I have no idea what CDC is doing biologically.

Student 2: We can look at the internet.

Students did not exhibit other aspects of NOS as strongly. They were less forthcoming in validating relationships, distinguishing between evidence and opinion, and realizing that insights from discovery-driven exploration take time and perseverance. Regarding validation, the majority of students neither recognized nor validated their mistaken read-offs from the Data Matrix when the numerals were inconsistent with the heat map color coding. Mistakes in reading these numeric values were due to a usability problem in the design of the Data Matrix. In the Matrix, numeric values were written with 15 decimal points, far too many to fit within the width of a Data Matrix column. Exceedingly small values were written in scientific notation, with the exponent presented at the end of the 15 decimals (e.g., $-8.\text{xxxxxxxxxxxxxxxxx}^{-4}$). Unfortunately, students could not see the exponent at the end.

With only some of the numerals visible, students believed the value was exceedingly large (a -8 -fold change in the example). Only four of the 11 pairs questioned why such strong fold changes were represented in color as black, which stood for small, inconsequential values. Notably, these same attentive four pairs had the lowest error durations (Pairs 3, 7, 10, and 11).

Of these four pairs, only three actually cross-referenced the Data Matrix numbers with the hover-over information for the color-coded cell to verify the seeming inconsistency. Among the three pairs attentive enough to cross-reference the Data Matrix, only two subsequently widened the column to uncover the exponent. These two pairs realized that the value, in fact, was exceedingly small, just as the heat map cell color denoted, and they revised the values in their worksheets accordingly. This left two of the four skeptical pairs placing their trust in the (still mistakenly read) tabular data over the color coding. Notably, no students questioned the mistakenly read numerals based solely on the supposedly high numeric values themselves. Students did not seem to remember that they had been told at the start of the module that fold change values ranged only from -2 to 2 in this data set.

Reflecting another gap in NOS knowledge, some students did not tackle the cognitive dissonance that they experienced. These students experienced dissonance when they saw that some CDC genes in the down-regulated subgroup were, in fact, colored red ("high up-regulated"). They commented that this occurrence was counterintuitive. Yet none of them questioned it further, for example wondering how the data had been pre-processed into down- and up-regulated subgroupings, which would have been a NOS question to raise. In terms of other aspects of NOS, students at times were not confident that they could

differentiate between evidence and opinion. Several explicitly questioned whether they could prove their point or if their claims were just opinions. Additionally, many students lacked patience during exploration. They often were so eager to “make a discovery” that they accepted counterintuitive findings without skepticism (e.g., a -9 fold change). They seemed to feel that discovery was some sort of contest. For example, they said: “Oh wow. Look. A negative nine...Ha, here’s another one right there: 6.42.” No students voiced a need to be patient when interesting relationships were not quickly forthcoming. In fact—quite the opposite—some students explicitly said it caused stress.

Visual Literacy and Visual Thinking—As part of visual thinking, students manipulated data and views, using the following program operations.

- Interact with the Radar, Pan, Zoom—to keep oriented visually
- Search using Active nodes, Focus, Scan for CDC strongest—to visually highlight elements of interest
- Focus, Arrange workspace, Arrange heat map and labels—to selectively attend to certain items

Students’ performances with these operations have already been discussed above. As they used these and other operations, students verbally expressed visual thinking. They confirmed and clarified what they saw, made sense of and judged the relevance of visually displayed relationships, and, in a minority of cases, validated inconsistencies. All 11 pairs visually recognized patterns and hotspots in the heat map and confirmed what they saw. All 11 pairs of students exchanged thoughts to clarify the content they saw. Clarifying content was one of the most frequently performed coded thinking behaviors for all students. The instructions only prompted students to clarify content once—at the start of the module. But five pairs continued to do so later.

Another frequently performed but unprompted visual thinking behavior was clarifying the meaning of color. All pairs of students except for one pair discussed whether they could assume from the colors that expression was inhibited or activated—and how strongly. Students’ frequencies and durations in this visual thinking behavior varied (see Fig. 3). Yet their clarifications paid off. At the end of the module, 86 % reported in the satisfaction survey that they could understand the meaning of color-coded “hot spots” in the heat map.

Additionally, all students with high-quality tapes for the CDC task engaged in the unprompted behaviors of classifying and comparing color shades. This type of visual thinking is noted in other studies as being common to scientific inquiry but overlooked by the New Biology curricula (Goodwin 1997). Students focused on classifying nuanced differences in color shade while characterizing fold changes in displays of aggregated CDC genes. They called colors light green, dark green, slightly red, or somewhat black. Mirroring conversations reported in Goodwin (1997) on color discrimination by collaborating scientists, students’ exchanges commonly went as follows:

Student 1: That’s not red.

Student 2: No

Student 1: I mean it's kind of red but it's not really red.

Student 2: [pointing to another cell] That one's red.

Student 1: So what's this one? Nothing? Black?

Student 2: Yeah.

This pair was somewhat distinct in pointing to a nearby heat map cell to compare and verify nuanced shades of color for classifying purposes. Only four other pairs made and reached agreement through such comparisons. Three pairs of students voiced frustration with nuanced color shades. They explicitly noted that perceptual “background noise”—e.g., surrounding colors or cell outlining during hover over—caused a bias that they could not correct for.

The principal difficulty in visual thinking experienced by students was getting a handle on the large volume of data. Sixty-four percent of students in the satisfaction survey reported being overwhelmed by the volume of the heat map. In the observational data, several students specifically complained about the volume of the Data Matrix. They found it tedious to search for strong fold changes in the dense data table with lengthy decimal places. The table had very little white space and little typographic leading or kerning, all of which determine legibility and usability. While scanning the table for high and low values, one student said to his partner: “Hey, I’m not getting anything out of this. It’s kind of dead overload.” Another pair abandoned the Data Matrix altogether.

Statistical Reasoning—All pairs of students except one applied statistical or mathematical reasoning to interpret or judge the displayed data. In one instance, seven of the pairs modified the statistical cutoff for judging an expression change to be strong in accord with their growing familiarity with the data values. At the start, the professor specified ± 0.4 as a good cutoff but over time seven pairs recast the criteria to ± 0.3 to widen the scope of potentially interesting relationships, acknowledging at the same time that sensitivity would lessen.

Another instance of positive statistical reasoning related to viewing and understanding the summary statistics for erg and CDC aggregates. Positively, all pairs knew what most of the statistics were and how they were derived. An exception, detailed further in an example below, was quartiles. Despite students’ positive understanding, however, 64 % of the pairs puzzled over the values of the color coding for variance and standard deviation. This puzzlement highlighted a usability flaw in the tool. The tool automatically applied the range of fold change values from the data set (-2 to $+2$) to render the color coding for all summary statistics. This scale made sense to students for means and medians but (rightfully) not for variances and standard deviations.

As evidence of difficulties with statistical reasoning, only a small handful of students (28 %) applied their knowledge about what summary statistics stood for to draw inferences from relationships among the statistics about the spread of individual values within an aggregate. Visual memory loads may have played a role in this omission. Students had to hold in mind at once numerous color shades for each statistic to draw relationships among them.

Generally, no more than three items at once can be held in visual memory (Bae and Flombaum 2013).

One pair of students experienced additional confounded problems related to summary statistics. As they viewed the seven different statistics, they changed color coding parameters so that slight differences in shade would become more pronounced. They set the parameters at -1 and 1 ; at -0.5 and 0.5 ; and finally at -0.2 and 0.2 . When the pair reviewed the colors they had recorded in the worksheet for each statistic, they concluded that “the first quartile is the only one that looks somewhat green so it must represent the group’s lowest value.” This pair did not acknowledge that, by definition, “minimum” should be the lowest statistical measure, not the first quartile. Instead, they trusted the colors that they had recorded: Minimum was black, and the first quartile was green (a lower value). The problem, which they never realized, was that they had read off the black minimum color with one range of color coding parameters and the green first quartile color with another range.

Higher-Order Thinking—Students were well-versed in the lower-level thinking skills on which higher level thinking for visual analytics rested. For example, students were facile in reading off numeric values and visual coding. They regularly brought in prior knowledge to enrich their understanding of views (e.g., “It’s an autophagy regulator. I used it in a presentation yesterday”). They used their own words and paraphrased to functionally describe relationships, regulatory events and underlying concepts (“It’s a protein that catalyzes demethylation in the ergosterol synthesis pathway. Okay, so if it’s part of the pathway it makes sense that mutating it [as a treatment] would really change things.”).

Fewer pairs revealed higher-order thinking. Verbal and written data provide evidence of the occurrence of this thinking. Some verbal instances are described in Table 5.

As mentioned, the third task focused specifically on higher-order thinking. Its omission diminished the variety of higher-order thinking that we could analyze. The second task nonetheless gave evidence of various types of analytic reasoning in the written explanations that students produced. Written explanations of the biology behind a strong treatment-CDC transcript fold change fell into five categories that were based on the number of factors students considered and the extent to which they related factors to contextual events or conditions. Descriptions of all five categories with sample passages as follows. The first two categories below revealed higher-order, multi-factor, situational thinking. Four pairs demonstrated such thinking. For the other pairs, such reasoning was missing.

- Explains possible intermediaries in the expression change process ($N = 3$ pairs)
cdc7 might be upregulated because Cycloheximide (drug) inhibits an inhibitor of cdc7.
- Explains expression change in terms of the functional traits of the treatment *and* gene ($N = 1$ pair)
cdc33 [which is down regulated] is a translation inhibition factor. Tunicamycin is a glycosylation inhibiting drug. The drug is used to induce unfolded protein response.

With it there would be no need for a translation inhibiting factor because translation can't occur.

- Explains expression change in terms of a functional trait of either the treatment *or* gene ($N = 3$ pairs)

For *cdc19* the drug *med2* could be blocking the kinase receptor site and thus blocking *cdc19*.

CDC33 is a translation initiation factor so inhibiting it with Tunacamycin will result in down regulation.

- Explains expression change tautologically (in terms of its numeric value) ($N = 1$ pair)

These drugs down regulate the transcription of these genes because the values we got are lower than zero.

- Explains by naming functional attributes but not their role in the expression change ($N = 1$ pair)

ccdc33 binds to mRNA cap protein and translation inhibition factor. Tunacamycin is an antifungal drug.

Time spent examining additional information for explanations seems to have influenced students' reasoning modes and the number of factors they considered. The data coded "Examining additional information" reveal how long each pair spent working on its respective explanation about CDC gene. Except for Pair 5, the longer a pair spent, the stronger their explanations and higher-order reasoning were (see Fig. 4). Interestingly, little correlation seems to exist between level of thinking in explanations and time spent on error (Fig. 5).

Omissions occurred. In addition to those already discussed—the third task; missed opportunities for multi-factor, situational thinking; and a failure to think through inconsistencies between the mistakenly read exponential values in the Data Matrix and the color coding—students did not synthesize information. They read external information to construct explanations but their reading did not lead to synthesis. Even though a majority accessed more than one source of information, the sources were typically expository—e.g., Wikipedia or public databases—providing little insight into "debates in the field." No pairs read research abstracts, articles or reviews.

Question 4: Usability. In What Ways, if Any, are omissions and Difficulties Tied to Tool Usability?

Usability problems detracting from student performance fell into two categories: (1) classic ease-of-use issues; and (2) support for comprehension and an empowering of extended thinking.

Most of the errors that students experienced were due to ease-of-use problems in operating the program, screen clutter or both (specified in Fig. 6). An additional usability problem that did not cause errors but that did cause slow and tedious performance was panning, which

dissatisfied 68 % of the students in the satisfaction survey. The only way to pan was to repeatedly drag in small steps the 6000 column-wide heat map.

Following the field study, the developers of CoolMap implemented improvements associated with these ease-of-use problems. For example, they designed drag and drop tabs and windows for easy workspace arrangement. They developed auto-position to bring the focus to the location at which users added aggregates to the heat map from the Ontology tab. They added shortcut keys for quick panning and dynamically linked the Radar and heat map.

The second category of usability—support for comprehension and empowered thinking—largely did not produce errors. But aspects of the program design did obstruct accuracy, burden cognitive load, and keep students from extending their thinking. Some instances have been discussed already, such as students' difficulties and subsequent inaccuracies in perceptually and cognitively processing the lengthy numerals in the Data Matrix and experiences of perceptual bias from surrounding colors. In post-study improvements, developers, respectively, changed numerical length to two decimal points and followed color theory guidelines reduce bias from adjacent hues and luminance. Another obstruction to comprehension related to the usability flaw of solely using color coding to represent summary statistics in aggregates. Most likely, holding in mind slight shade distinctions across sequentially viewed summary statistics burdened students' visual memories (Marshall and Bays 2013). Possibly, this burden hampered them from inferring the spread of values of aggregate members. In this case, aggregation was a positive means for decreasing volume, but it had the unintended consequence of increasing students' mental loads for making inferences. As usability studies show, assuring relevant data in displays is more important than decreasing volume if extraneous data for task purposes remains. As an alternate design, a program might pre-calculate and visually flag aggregated relationships in which values center around the mean or have outliers at the extremes, thereby giving visual cues for task-specific purposes. These automatic calculations would signal distributions of aggregate members' values without students having to do the mental processing.

Limitations

Many limitations exist for findings from the field study, and largely they trace to time constraints. As Meyer et al. (2013) note, time often is found to be an unexpected barrier to inquiry-based instruction aimed at higher-order learning. Retrospectively, two and a half weeks was not enough time for students to become familiar with the data and species, read outside sources from many perspectives, and complete the third task. Consequently, we do not have findings on many of the important processes involved in building higher-order explanations. Additionally, due to students' unfamiliarity with the data and species, we cannot confidently assess the core disciplinary concepts and cross-disciplinary concepts that they had or needed to have. That is, it is possible that students had more knowledge and understanding of multifactor biological relationships in context than shown—just not for these unfamiliar data and species. Correspondingly, they also may have demonstrated stronger NOS knowledge. Time constraints also affected our collection of data. In an effort to devote as much of the two and a half weeks as possible to task performance, we did not take the time to collect data on students' prior domain knowledge, visual literacy, statistical

reasoning or attitudes toward science. As a result, we were not able to associate performance with such traits, nor were we able to tape record whole class discussions between tasks. Rather during the discussion times, the research team had to render each pair's screen capture file quickly into exportable .mp4's so that Camtasia would then be ready on the pair's laptop to record the pair's next subtasks. Finally, in terms of tool improvements for usability, the fixes that CoolMap implemented follow usability principles and visual science theory but have not yet been formally tested for user performance.

Discussion

Our goal was to determine the feasibility of having undergraduate biology majors conduct the same inquiries as practicing scientists with interactive data visualizations. We focused on the NIH-funded heat map application, CoolMap, because it is a professional level, in-progress tool that we could immediately improve based on findings. Using observational and survey data, we examined students' efficiencies in their software-supported inquiries. We also uncovered various strengths and weaknesses in students' knowledge and skills that were relevant to visual analytics, focusing specifically on domain knowledge, a grasp of the nature of science, visual literacy and thinking, statistical reasoning, and higher-order thinking.

Students' performances and survey responses give evidence of feasibility. The biology majors brought relevant visual analytic knowledge and skill to bear to interactively uncover relationships of interest in the large data set. A majority of students exhibited fairly similar efficiencies in operating the program, and 80 per cent or more of the students completed all worksheet questions but one. In survey responses, large majorities indicated enjoyment in working with the tool and positive assessments of many program operations. Additionally, in terms of unifying thinking with perception and action, verbal exchanges and findings from coded thinking behaviors show that students facilely clarified the content and meaning of color to understand relationships. Without prompts, they mutually negotiated a shared understanding of color shadings and from them judged if fold changes were strong enough to be interesting. They found and biologically described patterns and relationships of interest. They grew increasingly familiar with the data set, and many students adjusted criteria and levels of confidence for judging "interesting relationships" accordingly. Additionally, some students—albeit a minority—positively applied higher-order thinking, inquiry-oriented NOS knowledge and attentiveness to inconsistencies, all of which are fundamental processes for sophisticated exploratory analysis. These behaviors suggest that the potential exists for sophisticated engagement in visual analytics. Moreover, students largely were technically savvy and unafraid of the new technology. They never questioned whether a professional-level visual analytic tool was outside their ken.

Evidence suggests a need to build on this solid foundation for feasibility. Specifically, many behaviors need to be strengthened that were associated with students' difficulties in visual thinking, statistical reasoning, nature of science knowledge and higher-order thinking. Technologically, this strengthening requires going beyond the usability fixes discussed earlier to include additional improvements. Instructionally, this strengthening requires

attention to several lessons learned informally from this study, as detailed below after the technology designs.

Technology Designs

Our field study findings underscore a need for tool improvements that we reviewed earlier from other studies on data visualizations for biology classrooms. Tools need to provide multiple perspectives, pre-defined calculations related to statistical reasoning and data transformations, and link-outs to information. These improvements could help students process aggregates visually, draw statistical implications from related summary values, and take diverse perspectives in their outside reading. We suggest the following improvements:

- *Built-in custom statistics combined with perceptual coding/flags.* In addition to visually flagging the pre-calculated statistics for aggregates in the heat map display (as discussed in the usability section above), certain numeric values in the Data Matrix also could be cued. For example, CoolMap could offer the option to highlight outlier values for a particular row or column.
- *“Smart” link-outs to PubMed.* Students’ complex thinking in explanations was tied to the time they spent reading outside sources. To facilitate doing this reading, a program could let students link from any given heat map cell (for individual level data) to PubMed. A click on the link would bring students to PubMed results from automatic queries on the variables labeled for that cell. This link out feature could get students to read relevant PubMed articles for different points of views about the targeted variables and events.
- *Annotation capabilities.* User-created annotations could aid students in remembering information and in extending their thoughts. In the erg task, students scanned the heat map for highest and lowest fold changes across all treatments, and in this scanning they often doubled back to look at a cell twice, three times, or more. Some students also compared cells across the heat map and compared the same cells than once. Students repeatedly read the hover-over information on the cell of interest without moving their cumulative sensemaking ahead. Annotated comments on earlier references to a cell would have enabled students to move earlier ideas forward in revisited cells (Quintana et al. 2004). Students also might annotate similarities between relationships, which could help them consider multiple factors. Additionally, they might annotate and better remember the color coding parameters under which they found a relationship to be “interesting.”

Instructional Approaches

Students’ performances suggest a need to strengthen, in particular, the NOS knowledge and higher-order thinking that visual analytics calls for. Though we did not formally analyze effects of instructional methods, students’ performances helped us recognize several instructional approaches that may have enhanced students’ NOS stances and higher-order thinking. Largely, these lessons learned are known in the educational research literature. But practically—given the demands of semester-long undergraduate courses—they often pose hard trade-offs for instructors, who often have to balance time on courses with time on

research and who have to balance breadth and depth of content in their courses. Our findings suggest that, though demanding, the following instructional approaches are important to visual analytic success.

Devote Ample Time to Running a Module

Students needed more than a total of seven and a half hours in a visual analytic module to construct multifactor explanations, to synthesize information, to interpret relationships from different perspectives, to generate defensible claims, and to distinguish novel from known relationships. Without more time in our field cases, the complexity of analysis and higher-order processes that students were able to conduct fell far short of the *exploratory and explanatory* analyses that scientists actually conduct with data visualization tools. Researchers in biology education contend that the “cycle of learning involved in problem-based inquiry can take many weeks” and suggest that curricula-targeting technology-supported, higher-order analysis should extend anywhere from three to eight weeks (DiCarlo 2006, 468). Understandably, devoting more time to a hands-on module results in having less time for other parts of the curriculum. One way to negotiate this tension may be for instructors to prepare students in the relevant NOS knowledge and higher-order thinking throughout the term for ready transfer to an upcoming case-based module.

Prepare Students in Relevant Processes Before and During the Module

Our field study results reinforce findings from other educational research on technology-enabled inquiries by suggesting teacher modeling may be an effective way to prepare students in relevant NOS knowledge and higher-order thinking for visual analytics (Songer et al. 2009). During modeling, teachers or other experts could engage students in many of the processes, make strategies explicit, revisit and practice the inquiry multiple times, and encourage transfer of understanding. For example, several modeling sessions throughout the semester could demonstrate the NOS stance of patience for discovering meaningful findings. The sessions could show that inductive reasoning and uncertainty are not comfort areas for anyone—even experts; but nonetheless novel insights depend on them. Modeling could also show that for counterintuitive read-offs (e.g., high expression of a gene in a down-regulated aggregate), analysts need to pose questions to resolve the dissonance, e.g., about the provenance or preprocessing of the data. In modeling sessions, a teacher or expert also could enact realistic explorations of tangents, thereby helping students to see the importance of this NOS activity. During modeling, students could see the teacher or expert successively attend to select patterns and cues and could begin to assimilate the tacit visual analytic process of being primed to look at certain configurations as potentially interesting. Selective attention is a transferable skill to all visual analytic cases (Lobata et al. 2012). In addition, a teacher or expert could run several preparatory modeling sessions throughout the course to help students compare thinking and reasoning across cases. Comparatively, teachers could help students examine the situated nature of many analytical decisions, instances that justify tangential investigation, and evidence is sufficient to warrant claims or inferences.

Design Cases for Performance Success

As mentioned, outcomes of the field study clearly show that data and species in cases need to be familiar to students. Outcomes also reveal two less obvious design considerations: (1) The structuring of learning progressions and (2) the balance between directed and undirected written instructions for higher-order thinking. Both issues are under-researched areas for inquiry-driven cases and higher-order thinking in undergraduate biology.

In the field study, it is unclear whether the structure of the learning progression was best for undergraduate biology majors. As noted, it conventionally moved students from lesser to greater conceptual and cognitive complexity. We adapted it from studies aimed at middle school biology students. Recently educational researchers have begun to question whether conventional lesser-to-greater progressions are best for college students (McElhinny et al. 2014). Researchers claim that because undergraduates already have a good deal of domain knowledge, using “inverted progressions” might better develop their capabilities. An inverted progression focuses on complex concepts at the start so as to spark students to tap into their prior or intuitive knowledge and mental models. Then, it moves to the conceptual building blocks and relationships underlying the complexity (McElhinny et al. 2014). Our field study does not provide evidence favoring one progression structure over the other. Qualitatively, students demonstrated shortcomings in applying a greater degree of higher-order thinking and NOS knowledge but many factors may have contributed to this outcomes. Such factors include but are not limited to the selected learning progression, the time constraints prohibiting progress to the third task, unfamiliarity with the data set and species, tool usability problems, and an imbalance between structured and open-ended instruction. Effects of these factors need to be systematically examined in future studies. By establishing the feasibility of incorporating visual analytics in undergraduate coursework, our field study lays the groundwork for such research.

The second design consideration—optimally apportioning directedness and open-endedness in written instructions—is a perennial challenge in field research. On the one hand, a great deal of prompting and directedness keeps students commonly “on task” for the sake of comparing student performance. On the other hand, pre-defining too many task actions can inhibit students’ inclinations to explore and patiently pursue leads without expecting imminent “solutions” (Wang et al. 2011). Our case may have had too much prompting. Positively, we found consistencies in students’ performance times, but their insights and explanations were limited. We believe that achieving an optimal balance between directedness and open-endedness is tightly coupled with providing students with clearly articulated case goals and rubrics that show how students’ analytical processes and outcomes will be assessed. For goals, students in our field study turned often to the written instructions to guide their next moves. For rubrics, had the instructions included them, it may have been less necessary to specify and bound actions so closely. Students instead could have set their own paths for achieving expected processes and outcomes.

Conclusions

This study assessed the feasibility of undergraduate biology majors using a professional-level interactive heat map tool to explore and explain functional effects of sequence variants.

Findings show that it is feasible but that usability designs need to be in place to consolidate this feasibility. Our study is a first step in recognizing that undergraduate biology majors likely can conduct the visual analytic work of practicing scientists with the same tools. Currently, the incorporation of visual analytics into undergraduate biology is in its infancy. Yet rapid growth is imminent. Life scientists increasingly conduct in-depth analysis with data visualizations to generate hypotheses. For professional training, it is an opportune time to integrate these applications into college classrooms. It is also an opportune time to use such experiences to improve the usability of data visualizations. Undergraduate users can uncover usability improvements that can benefit scientists and students alike. Subsequent versions of CoolMap have incorporated many of the designs suggested by field study findings. Our field study also suggests instructional considerations to enhance higher order and NOS thinking in case-based modules. Overall, to advance the emerging area of visual analytics for biology majors, more empirical studies and reports need to be shared, and new collaborations are needed between science educators and researchers, data visualization developers, and user experience and usability experts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by an NLM grant from the NIH, #1R01LM009812-01A2. We also would like to thank Juanita Lyons for her valuable contributions to the data analysis and Jean Song and Andy Lin for their help during the classroom module.

References

- AAAS. Vision and change in undergraduate education: final report. AAAS; Washington: 2011.
- Anderson, CW.; Bauerle, C.; DePass, A.; Donovan, S.; Drew, S.; Ebert-May, D.; Gross, L.; Hoskins, SG.; Labov, J.; Lopatto, D.; Lynn, D.; McClatchey, W.; Varma-Nelson, P.; O'Connor, C.; Pelaez, P.; Poston, M.; Singer, S.; Tanner, K.; Wessner, D.; White, H.; Withers, M.; Wood, W.; Wubah, D. Vision and change in undergraduate biology education: a call to action. AAAS; Washington: 2012. Accessed 5 Aug 2015 at: <http://visionandchange.org/finalreport/>
- Bae GY, Flombaum JI. Two items remembered as precisely as one: how integral features can improve visual working memory. *Psychol Sci*. 2013; 24(10):2038–2047. [PubMed: 23938276]
- Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huether CA. Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genet Educ*. 2008; 178(1):15–22.
- Colon-Berlingeri M, Burrowes P. Teaching biology through statistics: application of statistical methods in genetics and zoology course. *CBE-Life Sci Educ*. 2011; 10:259–267. [PubMed: 21885822]
- Creswell, J.; Clark, V. Designing and conducting mixed methods research. 2nd. SAGE Publications; Thousand Oaks: 2010.
- Crow A, Dirks C, Wenderoth MP. Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE-Life Sci Educ*. 2008; 7:368–381. [PubMed: 19047424]
- Gehlenborg N, O'Donoghue SI, Baliga N, Goesmann A, Hibbs MA, Kitano H, et al. Visualization of omics data for systems biology. *Nat Methods*. 2010; 7(3):S56–S68. [PubMed: 20195258]
- Goodwin, C. The blackness of black: color categories as situated practice. In: Resnick, L.; Saljo, R.; Pontecorvo, C.; Burge, B., editors. *Discourse, tools, and reasoning: essays on situated cognition*. Springer; Berlin: 1997. p. 111-140.

- Hughes T, Marton M, Jones A, et al. Functional discovery via a compendium of expression profiles. *Cell*. 2000; 102:109–126. [PubMed: 10929718]
- Iskopehi, RD. Visual analytics in biology curriculum network. AAAS vision and change in undergraduate biology education initiative. 2014. Accessed 7 Dec 2014 at: <http://visionandchange.org/page/3/?s=vision?and?change?a?call?to?action>
- Johnson, C.; Moorhead, R.; Munzer, T.; Pfister, H.; Pheingans, P.; Yoo, T. NIH-NSF visualization research challenges report. 2006. Accessed 5 Aug 2015 at: <http://tab.computer.org/vgtc/vrc/NIH-NSF-VRC-Report-Final.pdf>
- Keim, D.; Kohlhammer, J.; Ellis, G.; Mansmann, F., editors. Mastering the information age solving problems with visual analytics. Eurographics Association; Goslar: 2010.
- Kumar A. Teaching systems biology: an active learning approach. *Cell Biol Educ*. 2005; 4:323–329. [PubMed: 16341259]
- Lavie, N. Selective attention and cognitive control: dissociating attentional functions through different types of load. In: Monsell, S.; Drive, J., editors. Control of cognitive processes: attention and performance XVIII. MIT Press; Cambridge: 2000. p. 175-194.
- Lee SWL, Tsai CC. Technology-supported learning in secondary and undergraduate biological education: observations from literature review. *J Sci Educ Technol*. 2013; 22:226–233.
- Lemons P, Lemons JD. Questions for assessing higher order cognitive skills: it's not just Bloom's. *CBE-Life Sci Educ*. 2013; 12:47–58. [PubMed: 23463228]
- Lengler, R. Proceedings of the 10th annual conference on information visualization. London, England: 2006. Identifying the competencies of 'visual literacy': a prerequisite for knowledge visualization; p. 232-236.
- Levy D. How dynamic visualization technology can support molecular reasoning. *J Sci Educ Technol*. 2013; 22:702–717.
- Liu Z, Nersessian N, Stasko J. Distributed cognition as a theoretical framework for information visualization. *IEEE Trans Visual Comput Graphics*. 2008; 14(6):1173–1180.
- Lobata J, Rhodehamel B, Hohensee C. "Noticing" as an alternative transfer of learning process. *J Learn Sci*. 2012; 21(3):433–482.
- Marshall L, Bays PM. Obligatory encoding of task-irrelevant features depletes working memory resources. *J Vision*. 2013; 13(2):21.
- McElhinny T, Dougherty M, Bowling B, Libarkin J. The status of genetics curriculum in higher education in the United States: goals and assessment. *Sci Educ*. 2014; 23:445–464.
- Meyer DZ, Meyer AA, Nabb KA, Connell MG, Avery LM. A theoretical and empirical exploration of intrinsic problems in designing inquiry activities. *Res Sci Educ*. 2013; 43:57–76.
- National Research Council (NRC). BIO2010: transforming undergraduate education for future research biologists. National Academies Press; Washington: 2003.
- National Research Council (NRC). A new biology for the twenty-first century: ensuring the United States leads the coming biology revolution. National Academies Press; Washington: 2009.
- National Research Council (NRC). Discipline-based education research: understanding and improving learning in undergraduate science and engineering. The National Academies Press; Washington: 2010.
- Ngyuen N, Nelson A, Wilson T. Computer visualizations: factors that influence spatial anatomy comprehension. *Anat Sci Educ*. 2012; 5:98–108. [PubMed: 22232125]
- Nielsen, J. Response time: the three important limits. 1993. Accessed 11 Dec 2014 at: <http://www.nngroup.com/articles/response-times-3-important-limits/>
- Oviatt S, Cohen A. Toward high-performance communications interfaces for science problem solving. *J Sci Educ Technol*. 2010; 19:515–531.
- Peer A, Shneiderman B. Integrating statistics and visualization: case studies of gaining clarity during exploratory analysis. Proceedings of the SIGCHI conference on human factors in computing systems (CHI'08). 2008:265–274.
- Peters VL, Songer NB. Evaluating the usability of a professional modeling tool repurposed for middle school learning. *J Sci Educ Technol*. 2013; 22(5):681–696.

- Quintana C, Reiser BJ, Davis EA, Krajcik J, Fretz E, Duncan RG, et al. A scaffolding design framework for software to support science inquiry. *J Learn Sci.* 2004; 13:337–386.
- Sadler T, McKinney L. Scientific research for undergraduate students: a review of the literature. *J Coll Sci Teach.* 2010; 39(5):43–49.
- Sadler T, McKinney L, Hogan K. Exploring a process view of students' knowledge about the nature of science. *Sci Educ.* 2000; 84:51–70.
- Saraiya P, North C, Duca K. Visualizing biological pathways: requirements analysis, systems evaluation, and research agenda. *Inf Vis.* 2005a; 4:191–205.
- Saraiya P, North C, Duca K. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Trans Visual Comput Graphics.* 2005b; 11(4):443–457.
- Scholtz J. Developing guidelines for assessing visual analytics environments. *J Inf Vis.* 2011; 10(3): 212–231.
- Shubert C, Ceraj I, Riley J. Bringing research tools into the classroom. *J Comput Math Sci Teach.* 2009; 28(4):405–421.
- Smolinski T. Computer literacy for life sciences: helping the digital-era biology undergraduates face today's research. *CBE Life Sci Educ.* 2010; 9:357–363. [PubMed: 20810969]
- Songer N, Kelcey B, Gotwals A. How and when does complex reasoning occur? *J Res Sci Teach.* 2009; 46(6):610–631.
- Spiro, RJ.; Feltovich, PJ.; Jacobson, MJ.; Coulson, RL. Cognitive flexibility, constructivism and hypertext: random access instruction for advanced knowledge acquisition in ill-structured domains. In: Duffy, T.; Jonassen, D., editors. *Constructivism and the technology of Instruction.* Erlbaum; Hillsdale: 1992.
- Su, G. Unpublished doctoral dissertation. University of Michigan; Ann Arbor, Michigan: 2013. Omics data exploration: across scales and dimensions.
- Trujillo C, Cooper M, Klymkowsky MW. Using graph-based assessments within Socratic tutorials to reveal and refine students' analytical thinking about molecular networks. *Biochem Mol Biol Educ.* 2012; 40(2):100–107. [PubMed: 22419590]
- Wang, X.; Dou, D.; Butkiewicz, T.; Bier, E.; Ribarsky, W. A two stage framework for designing visual analytics systems in organizational environments. *Proceedings of the IEEE visual analytics science and technology conference*; Providence, RI. 2011.
- Wefer S, Sheppard K. Bioinformatics in high school biology curricula: a study of state science standards. *CBE-Life Sci Educ.* 2008; 7:155–162. [PubMed: 18316818]

Fig. 1. CoolMap design. **a.** Initial screen for reading in data. **b.** Yeast data set displayed showing the heat map in the center canvas with treatments labeled in the columns and transcripts labeled in the rows. Hover-over information shows the fold change value in expression for a treatment-transcript relationship. Widgets surrounding the heat map include (clockwise from the *top left corner*): Search (called Active Node); Zoom control tab; tabbed widgets for the Data Matrix (table of numeric fold change values) and the Ontology table (for adding aggregates to the heat map); Color coding; and Aggregator pull down list to select a summary statistic to display for aggregates. The Radar (context view) is not displayed and is an alternate tab to Color Coding for display. **c** Zoomed-in view of color-coded fold changes on a scale of -2 to +2 (*green* = low/negative and *red* = high/positive)

Fig. 2.

For 62 % or more of the pairs, the percent each pair spent, respectively, in a program operation

Fig. 3.
Total minutes spent in error (*blue*) and resolving error (*red*), by pair (Color figure online)

Fig. 4.
Frequency and duration in seconds for clarifying the meaning of color, per pair (Color figure online)

Fig. 5.
Pairs' time (in minutes) on additional information color coded by explanatory approaches
(Color figure online)

Fig. 6. Errors related to usability issues and pairs who experienced each one, colored by error type (Color figure online)

Table 1

Erg Gene task and its intellectual demands

Subtasks	Intellectual demands	Prompted
Apply criteria for judgment	Apply a cutoff value (pre-defined in the instructions) to identify “strong” fold change values	✓
	Adjust the criterion value as needed based on evolving familiarity with the data	
Identify, judge, explain strong relationships	Identify visual and numeric evidence of strong up/down expression change in an aggregate and at the individual member level	✓
	Use evidence to support judgments of strength	
	Biologically explain a strong treatment-expression relationship	✓
	Consider and speculate about the role of ergosterol biosynthesis in the strong relationship	
Validate	Validate consistencies in table-displayed values and color coding	
Extend considerations for additional meaning	Look for instances of similar or opposite responses of different genes to a treatment	
	Tangentially, seek information to explain similar/opposite responses	

Table 2

CDC Gene Task and its intellectual demands

Subtasks	Intellectual demands	Prompted
Infer spreads of values	Change and read seven summary values for a treatment-sub-group cell	✓
	Draw relationships between summary values to infer the spread of values across members of the sub-group	✓
Identify and judge strong relationships	Repeat the same analytical procedures as in Task 1	✓
	Understand and judge relationships according to the two integrated dimensions of cell and treatment functions	
Construct explanations	Search for and use relevant outside sources	✓
	Synthesize information from sources into explanations	

Table 3

Hierarchical clusters of gene task and its intellectual demands

Subtasks	Intellectual demands	Prompted
Identify, judge and explain strong relationships	Identify clusters of strong effects that have 1 + uncharacterized gene	✓
	Identify functions of known genes in the targeted cluster	✓
	Speculate about how the uncharacterized gene may associated with these genes in relation to the treatments affecting the cluster	
Construct explanations	Project investigative processes and content for next steps in finding possible connections between genes in the cluster and the treatments	✓
Evaluate	Comparatively evaluate the strengths and limits of clustering and functional aggregation in heat map to find strong relationships	✓

Table 4

Instructional agenda

Day	Activity
1	Class discussion of the research article providing the case and data set (Hughes et al., 2000)
2	40-min training on CoolMap; 35-min CoolMap-supported, case-based tasks in self-selected pairs + class discussion in between subtasks (erg task)
3	CoolMap-supported, case-based tasks in pairs + class discussion (erg task)
4	CoolMap-supported, case-based tasks in pairs + class discussion (CDC task)
5	Half day CoolMap-supported, case-based tasks + class discussion (CDC task); half day wrap up

Table 5

Demonstrated higher-order thinking

# Pairs	Higher-order thinking and examples
6	<p>Reasons why certain treatments activate erg gene expression and the role of ergosterol synthesis Exemplified in worksheet entries: Itraconazole is an antifungal drug that affects the cell wall ...and may be affecting repressors of the transcription of erg genes</p> <p>Itraconazole is an antifungal drug that affects the cell wall ...and may be affecting repressors of the transcription of erg genes</p> <p>The HMG2 gene is responsible in the cholesterol biosynthesis pathway so greater expression may occur in response to the deletion of HMG2</p>
4	<p>Understand/judge expression change based on integrated dimensions of a cell and treatment</p> <p>Exemplified in conversation:</p> <p><i>Student 1:</i> [Reading from outside source] CDC33 [gene] is an mRNA cap binding protein and translation inhibitor factor. What does tunicamycin [treatment] do?</p> <p><i>Student 2:</i> [Reading from Wikipedia] Tunicamycin inhibits....the glycoprotein synthesis. That makes sense with the cell wall thing. It blocks synthesis of all of these genes [cell division cycle] and causes cell cycle arrest in G1</p> <p><i>Student 1:</i> This is about translation. So the cell is....</p> <p><i>Student 2:</i> I guess it's being translated then. I guess Tunicamycin would bump down translation. [Reads]</p> <p>Yes, see, here. It says here Tunicamycin induces unfolded protein response</p> <p><i>Student 1:</i> But that's not like preventing sequences from being translated from the mRNA. That's like preventing folding and causing protein degradation. That's not the same. That's like the other side of it</p>
2	<p>Address cognitive dissonance about a down-regulated subgroup having a high positive fold change</p> <p>Exemplified in conversation:</p> <p><i>Student 1:</i> Why is the median nothing [i.e., black]. Aren't these down -regulated [and presumably green]?</p> <p><i>Student 2:</i> Where was that one red one? Is it enough to offset all the others in the group?</p> <p><i>Student 1:</i> [Hovers over the red one] It's .158. It's really red</p> <p><i>Student 2:</i> But it isn't the strongest down-regulated. Like it's upregulated right? It's red</p> <p><i>Student 1:</i> But it's still in the down-regulated group. [Goes down the column to look at values for the other genes in the group]. This one's higher, -0.179, and negative. All right technically it's the strongest</p>