

Enhancing Genomic Learning through Tabletop Interaction

Orit Shaer¹, Megan Strait², Consuelo Valdes¹, Taili Feng¹, Michael Lintz³, and Heidi Wang¹

¹Wellesley College, 106 Central St, Wellesley, MA, USA, 02481

²Tufts University, 419 Boston Ave, Medford, MA, USA, 02155

³Franklin W Olin College of Engineering, 1000 Olin Way, Needham, MA, USA 02492

ABSTRACT

We present G-nome Surfer 2.0, a tabletop interface for fostering inquiry-based learning of genomics. We conducted an experimental study with 48 participants that compared students' learning of genomic concepts using existing bioinformatics tools and using two alternative implementations of G-nome Surfer: a collaborative multi-mouse GUI and a tabletop interface. Our findings indicate that G-nome Surfer improves students' performance, reduces workload, and increases enjoyment. The comparison of tabletop and multi-mouse implementations further shows that the tabletop condition results in four educational benefits: 1) increasing physical participation, 2) encouraging reflection, 3) fostering effective collaboration, and 4) facilitating more intuitive interaction.

Author Keywords

Collaborative learning, multi-touch, tabletop interaction

ACM Classification Keywords

H5.2 Information Interfaces and Presentation: *User Interfaces*

General Terms

Design, Human factors

INTRODUCTION

The field of genomics informs both high-impact public policy decisions and personal health choices in areas such as cell biology, transgenic agriculture, and cancer research. Thus, it is viewed as essential to train undergraduate students to understand biologically meaningful insights from genomic data [32]. The study of genomic data depends on computational tools such as genome browsers and basic local alignment search tools (BLAST). However, existing state-of-the-art tools pose high threshold for information finding and do not support important aspects of college-level science learning such as inquiry-based high-level reasoning, the development of process knowledge, and collaboration [2, 17, 29, 31]. Motivated by the need for better computational support for inquiry-based learning of college-level genomics and by previous work illustrating the potential of tabletop interfaces to facilitate collaboration

[1, 3, 7, 10, 12], active reading [18], and distributed cognition [20], we developed G-nome Surfer 2.0, a tabletop interface for fostering inquiry-based learning of genomics. G-nome Surfer 2.0 draws upon G-nome Surfer 1.0 [29] that was designed to support research activities in genomics rather than students' learning. G-nome Surfer 2.0 introduces new and redesigned features that *explicitly support* collaborative inquiry-based learning as well as additional data sets and new architecture that leads to improvement in performance.



Figure 1, G-nome Surfer 2.0: Comparing ontology and gene expression information of different mouse genes.

In this paper, we describe the design of G-nome Surfer 2.0 (see Figure 1), which is a result of an iterative process in which we worked closely with biology and neuroscience instructors to identify requirements for supporting students' learning. We also present an experimental study with 48 biology students that explores the strengths and limitations of G-nome Surfer in supporting collaborative learning. To identify which of G-nome Surfer's strengths are products of its visual design and integrated workflow, and which are results of a multi-touch tabletop interface, we conducted an ecologically valid comparison between a traditional GUI consisting of state-of-the-art web-based bioinformatics tools and two alternative implementations of G-nome Surfer: a collaborative multi-mouse GUI and a tabletop interface. Since adoption barriers for tabletop interfaces are relatively high and are influenced by cost, size, and unclear

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

educational benefits [22], we aim to highlight the advantages of choosing tabletop interaction for collaborative learning in college settings. Our hypothesis is that the tabletop interface provides better support for collaborative learning both in terms of *efficiency* (the amount of mental effort required for learning) and *effectiveness* (the extent to which learning goals are obtained as a result of productive discussion). To measure the efficiency and effectiveness of collaborative learning we considered performance, attitude and workload, as well as verbal and physical aspects of interaction.

The rest of the paper is organized as follows: the next section summarizes related work, following that, we discuss the three contributions of this paper. The first is a set of design requirements for supporting inquiry-based learning of genomics; the second is G-nome Surfer 2.0, a tabletop interface for collaborative learning. Finally, we present findings from a comparative study that highlights the strengths of G-nome Surfer 2.0 compared to existing state-of-the-art tools and identifies four educational benefits of tabletop interaction compared to multi-mouse GUI.

RELATED WORK

Reality-Based Interfaces for Science Education

A number of systems illustrate the potential of supporting science education through reality-based interaction [15]. Here we describe those most closely related to our work. Gillet et al. [5] created a tangible user interface for molecular biology that uses augmented reality technology to enhance 3D molecular models. Chemiearaum [6] and [4] are tangible user interfaces for chemistry education. Schkolne et al. [28] developed a tangible immersive interface for the design of DNA molecules. These systems highlight potential contributions to science education in areas closely tied to genomics but focus on objects that have an inherent physical structure. We are interested in cases where *abstract* information is represented. Involv [9] is a multi-touch tabletop interface for exploring the Encyclopedia of Life that shares our challenge of creating effective interaction techniques for massive data spaces. However, it targets museum settings, which differs from college classroom settings.

Interactive Surfaces and Collaborative Learning

Several studies have investigated the effects of different interactive surface parameters on collaborative work including the orientation of the display [23], table size [25] and input techniques [1,10]. However, relatively little work has explicitly examined the effects of interactive surface parameters on collaborative learning. Of the studies that have been directly investigating collaborative learning, most have focused on children [1,3,7]. Piper and Hollan [22] conducted a study with pairs of undergraduate students, comparing the affordance of tabletop displays and traditional paper handouts. Our study draws upon this body of work but compares tabletop interaction to multi-mouse GUI in the context of college-level learning.

Learning and Gestures

A substantial body of work indicates that our hands play an important role in learning and thinking [14, 24]. Various studies provide evidence for the importance of gestures in face-to-face communication of scientists [19, 33]. These studies indicate that deictic gestures allow scientists to discuss abstract concepts and describe visual displays. Based on this view that gestures are integral part of the collaborative learning process, we recorded and analyzed gesture data along with verbal and touch data.

G-NOME SURFER 2.0

G-nome Surfer 2.0 is a tabletop interface for fostering inquiry-based learning of genomics. It draws upon G-nome Surfer 1.0 [29] that was designed to facilitate collaborative research activity through integrated information workflow. G-nome Surfer 2.0 introduces new functionality for explicitly supporting collaborative inquiry-based learning including: a contextual help tool, glossary, pair-wise alignment, integrated presentation of anatomical and microarray data, and workflow visualization. In addition, we redesigned the information artifacts introduced in G-nome Surfer 1.0 (e.g. ontology, sequence data, and publications) to facilitate spatial organization of information, and integrated data sets from JAX Mice and Allen Brain Atlas that were not available from G-nome Surfer 1.0. Finally, we implemented architectural changes in G-nome Surfer 2.0 that improve its performance. Following, we describe design implications, interaction techniques, and the implementation of G-nome Surfer 2.0.

Investigating Inquiry-based Learning

To identify design implications for supporting collaborative learning of college-level genomics, we collected data from a series of recurring individual one-hour meetings with five instructors and from examining related teaching materials. We analyzed the data by identifying common themes and distilling design implications. Following, we discuss four themes that informed the design of G-nome Surfer 2.0.

Multiple Representations

Genomics learning requires the interpretation and interrelation of diverse visual representations including diagrams, micrographs, and notations. To interpret representations students are required to process elements such as symbols and conventions and relate them to relevant content knowledge. Some visual representations convey the same biological idea from different perspectives. For example, a protein sequence can be described using an amino acid or a nucleotide sequence. But since each bioinformatics tool utilizes particular conventions, students often have difficulties interrelating different representations. As one instructor describes, “Students get confused with the letter references to amino acids and need help distinguishing this sequence from a DNA sequence” [21]. Thus, to facilitate learning, there is a need to link visual elements to relevant content knowledge and to allow users to relate different representations.

Process Knowledge

According to the National Research Council [30], effective inquiry-based units are designed to teach science content as well as process. In the context of genomics, process knowledge includes an information workflow that links together several data sets, each being handled with a special bioinformatics tool. In instructional units that require students to synthesize large amounts of heterogeneous information the workflow is rarely linear and involves multiple intermediate data formats. Thus, it is often difficult to navigate, generalize, and repeat such information workflow. Therefore, there is a need for visualizing the information workflow so that the computations applied in each stage can be easily distinguished and repeated.

Reflection and Discussion

Research indicates that students' understanding of the nature of science is enhanced through reflection and discussion [30]. Thus, inquiry-based units typically involve students working in groups to solve an over-arching question that necessitates multi-stage investigation. Between stages, students are required to reflect, discuss, and make decisions. In genomics, computational tools play a key role in multiple stages of the investigation. Therefore, it is important that such tools facilitate communication and reflection. However, current bioinformatics tools utilize a traditional GUI that constrains gestural abilities and limits the opportunities for face-to-face communication [10]. This suggests that less constraining human-computer interaction styles such as tabletop interaction may be more effective for inquiry-based learning. Reflection can also be promoted by providing users with means for spatially manipulating and easily annotating information artifacts.

Heterogeneous Information

To provide insight into complex biological systems, inquiry-based units often deal with information at varying levels of abstraction. For example, to *classify* the source of a motor mutation in mice, students need to investigate information from whole chromosomes down to an individual nucleotide position. In addition, instruction units often require students to access and relate large amounts of *heterogeneous* information. For example, to *identify* a motor mutation in mice students are required to investigate behavioral, anatomical, and genomic data. To facilitate learning, computational tools need to provide means for seamless transition between different levels of biological information and for relating *heterogeneous* information.

Designing for Collaborative Learning

Based on these themes, we defined a set of design goals for explicitly supporting collaborative learning in college-level genomics: 1) promoting discussion and reflection, 2) providing means for relating information both horizontally (i.e. across different biological entities) and vertically (i.e. different biological levels of the same entity), 3) facilitating the construction of both content and process knowledge, and 4) highlighting the relationships between different forms of visual representations. Following, we describe G-

nome Surfer 2.0's interaction techniques designed to meet these goals. G-nome Surfer 2.0 was designed in an iterative process that involved consistent feedback from instructors.



Figure 2, Pair-wise alignment of RNA segments with annotations (left); Contextual help displaying reference information related to RNA (center); Workflow visualization shows operation performed on three genes (right).

Bridging Across Representations

We developed several new mechanisms to help users interrelate diverse representations. The *Contextual Help* tool provides references to specific visual structures within G-nome Surfer. Placement of a tangible flashlight upon a selected area displays a glossary definition that ties the visual representation to relevant content knowledge. The definition is supplemented with links to related terms and a scrollable alphabet index. For example, when the user places the flashlight upon an amino-acid sequence, a definition and an amino-acid symbol key appear. The information fades upon flashlight removal. G-nome Surfer provides three flashlights that can be used in parallel. We chose to represent the help tool with a tangible object in order to increase visibility and encourage users to discuss researched terms. The flashlight shape was chosen as a metaphor for “illuminating” an area.

The *pair-wise alignment* feature allows users to identify regions of similarity between sequences. When a user aligns one sequence on top of another, the two sequences “snap” together and a new window appears containing a matrix in which the two sequences are represented as rows so that similar characters are aligned in successive columns and areas of discrepancy are highlighted. This allows users to explore the relationship not only between sequences but also between alternative representations of the same sequence (e.g. DNA, RNA, amino-acid). Figure 2 shows the pair-wise alignment and contextual help features.

The *gene expression* visualization was functionally and visually redesigned based on users' feedback. This visualization displays an image of an organism with several regions highlighted across the body. When a region is

tapped, an additional panel slides out from the right side of the window, showing a detailed list of tissues related to that area. The expression information is coded using a monochromatic scheme. In an earlier version, this visualization used a red-green color scheme that is common in existing bioinformatics tools. However, we found that this color scheme often confuses users and results in numerous mistakes. Additionally, for selected areas of an organism's body (e.g. brain) where the construction of anatomical knowledge is important, the new visualization presents a corresponding anatomical micrograph. We designed this visualization to enable users to relate abstract expression representations to concrete anatomical entities. Figure 1 shows the gene expression feature.

Spatial Organization

To help users spatially relate information both horizontally (i.e. across different biological entities) and vertically (i.e. different biological levels of the same entity), each information artifact is labeled by gene as well as coded with a color. We utilize a new color-coding scheme that differentiates types of biological information. For example gene ontology is marked with a blue border, publications with grey border, and sequences with green border.

	Traditional GUI	Multi-mouse GUI	Tabletop
Display size	30"	30"	30"
Display orientation	Vertical	Vertical	Horizontal
Interaction style	WIMP, single user	Bimanual, multiple users	Multi-touch
Workflow	Distributed	Integrated	Integrated
Input	Indirect	Indirect	Direct
Tangible interaction	No	No	Yes

Table 1, Differences across experimental conditions.

Workflow Visualization

The workflow visualization records and displays all interactions performed upon G-nome Surfer (see Figure 2). Docked at the upper right-hand corner, this visualization can be accessed upon request with a simple drag. Actions are organized by gene and are displayed as color-coded blocks on a gene entry. The color-coding is consistent with the scheme used for differentiating information artifacts. When switching between multiple genes, new entries are added into the visualization chronologically. Users may view the entire workflow of a particular gene by expanding the corresponding entry. This visualization facilitates the construction of process knowledge by allowing users to record, and repeat a workflow.

Implementation

G-nome Surfer 2.0 draws upon the Microsoft Surface implementation of G-nome Surfer 1.0 [29]. However, while G-nome Surfer 1.0 relied on web services for retrieving data, in G-nome Surfer 2.0 we developed and integrated a

local database of RefSeq genes. We also implemented real-time sequence comparison and analysis using the NCBI BLAST+ tools suite. This implementation improves the performance of G-nome Surfer so that it is comparable with existing bioinformatics tools.



Figure 3, Comparing similarity search results on the three experimental conditions: traditional GUI with multiple Firefox windows displaying UCSC Gnome Browser, JAX Mice, Entrez Gene and NCBI Blast (left); a multi-mouse implementation of G-nome Surfer 2.0 running on the Microsoft Surface (center), physical interaction objects were replaced with buttons; G-nome Surfer 2.0 (right).

EXPERIMENT

To investigate G-nome Surfer's strengths and limitations in supporting collaborative learning, we conducted a between-subjects experiment with 48 undergraduate students comparing the system to both current state-of-the-art tools and to a collaborative multi-mouse GUI. In this study we examined the similarities and differences in terms of quantitative performance and qualitative behavior in 24 dyads that worked on an inquiry-based task. Our goal was to identify differences in both the efficiency and effectiveness of the collaborative learning process across experimental conditions.

To ensure the ecological validity of our comparison, we consulted with genomics instructors and decided to compare G-nome Surfer 2.0 *tabletop interface* with the computer setting currently used for teaching inquiry-based genomic courses, a single mouse *traditional GUI* consisting of a set of web-based-tools, as well as with an alternative prototype of G-nome Surfer 2.0 implemented as a collaborative *multi-mouse GUI* using the Microsoft Surface Simulator. In both G-nome Surfer conditions (tabletop and multi-mouse GUI) there is a minimal need for using the keyboard. The core functionality supported by all conditions includes: gene search, ontology information, publications search, sequence retrieval, gene expression, BLAST search, and pair-wise alignment. To keep the conditions comparable we removed the contextual help and workflow visualization from the G-nome Surfer's conditions, as existing tools do not offer similar functionality. Several other factors vary across these conditions. Table 1 summarizes the difference between the experimental conditions. We discuss possible effects of these differences in the Results and Discussion section. Figure 3 shows the apparatus of each condition.

Experimental Task

Research on collaborative learning shows that the more complex a task is (i.e. the higher the intrinsic cognitive load), the more effective it is for learners to collaborate

[13]. Thus, our goal was to select a task complex enough to foster effective collaborative learning while maintaining authenticity. To satisfy this requirement, we selected a task that mirrors an inquiry-based instructional unit from an intermediate undergraduate neuroscience course [21]. This unit was designed with clear and measurable learning outcomes and with a goal of incorporating students' reflection and discussion [32]. We selected to use a collaborative paired condition because in the course students work in pairs to complete this instructional unit.

Our experimental task consists of two subtasks: an investigation of the human cancer-related gene TP53, and an investigation of a motor mutation in mice. Each subtask consists of several activities including information retrieval, information interpretation, decision-making, and reflection. The first subtask, an investigation of the structure and expression of TP53, aims to familiarize the participants with the interface. In the second subtask, subjects are required to identify the molecular basis of an unclassified motor mutation in a mouse. To accomplish this goal, subjects must access and relate large amounts of heterogeneous genomic information. In both tasks, participants follow specific steps and record answers to questions that reflect biological insights as they work.

Measures of Collaborative Learning

To test the hypothesis that the tabletop condition provides a better support for collaborative learning both in terms of efficiency and effectiveness compared to the other conditions, we utilized a set of quantitative measures and qualitative indicators. Collaborative learning is considered *effective* if learning goals are obtained and the collaboration process include discussion and reflection upon the task, thus leading to deeper and meaningful learning [13]. In effective collaboration, group members must actively communicate with each other to demonstrate shared effort [13]. Collaborative learning is considered *efficient* if learning goals are obtained with the investment of less mental effort [13]. To measure *efficiency*, we recorded total session and task *time*, and time for each activity. We also used the NASA Task Load Index (TLX) [8] for measuring *subjective workload*, and discourse analysis to quantify utterances focused on *interaction syntax*.

To measure the *effectiveness* of the collaborative learning process, we looked at a collection of indicators including the obtainment of learning goals, participant attitudes, levels of participation, nature of discussion, collaboration styles, and problem solving strategies. To quantify the *obtainment of learning goals* in each session, each dyad was asked to answer a set of biology questions while working towards completing the task. This questionnaire is similar to the questionnaire used in the neuroscience course. We assigned a total score for each dyad based on the correctness of their answers. We measured *participant attitudes* using subjective measures of enjoyment and

confidence that were collected from each participant in a post-task questionnaire.

We also looked at both verbal and physical participation. First, we used transcripts of each session to calculate levels of *verbal participation* per participant. Then to compute a

Talk Category	Definition	Examples from transcripts
Insight	An individual observation about the data, a unit of discovery [27].	'It's a tumor suppressant.' 'I think it's not SOD1 cause we know the mice have bad motor coordination, and are small.' 'This looks exactly the same as that, they're probably related.'
Coordination	Utterances referring to turn taking, verbal shadowing [3], or task division.	'Do you want to write it or do you want me to?' 'Let me search for Rora then.' 'Wait, get to the other one.'
Brief response	Short responses to suggestions or moves.	'Ok', 'Sure', 'Cool', 'Oh.'
Syntax	Utterances referring to how to use certain features.	'How do I scroll?' 'Is that the exon menu?' 'Try dragging it out' 'How do we BLAST?'
Problem solving	Utterances such as planning, relevant questions or pointing out facts.	'What do we have to do next?' 'We're suppose to overlay the RNA and amino acid sequences' 'Let's find amino acid C'
Disengagement	Non-task related questions and comments.	'Your lab is awesome' 'That could be really hypnotic' 'So you're doing research?'

Table 2, Coding scheme for classifying talk categories.

measure of verbal participation per session, we calculated the mean number of utterances per minute for each dyad [7]. Second, we used video analysis to calculate levels of *physical participation*. Since gestures are a complementary modality to speech, we included 'offline gestures' - movements of the hands, and other body parts, along with touch events in our measure for physical participation. We calculated levels of physical participation per participant by summing the number of touch events and offline gestures. Then we calculated a measure for physical participation per session by summing the mean number of touch events per minute and the mean number of offline gestures per minute. To measure the relative contribution of individuals within each session, we used the *Gini Coefficient* as a measure of the *equality of participation* [7].

We were also interested in understanding the differences in the *nature of discussion* across the different conditions. We iteratively developed a coding scheme that classifies talk into six general categories. Table 2 defines each talk category and provides an example from the transcript for each type. Finally we used qualitative indicators including those introduced by Meier et al. [16] and by the CLM

framework [3] to analyze collaboration processes and problem solving strategies.

Participants

48 undergraduate students (32 female, 16 male, age range: 18 to 24) volunteered to participate in the study, making up 24 dyads. All of the participants had successfully completed at least one introductory Biology course. 56 percent of the participants were biology or neuroscience majors. 39 percent of the participants had some experience with bioinformatics tools. None of the participants have taken the Neuroscience course in which the instructional-unit used for our task was taught. In six dyads (two on each condition) participants were familiar with each other.

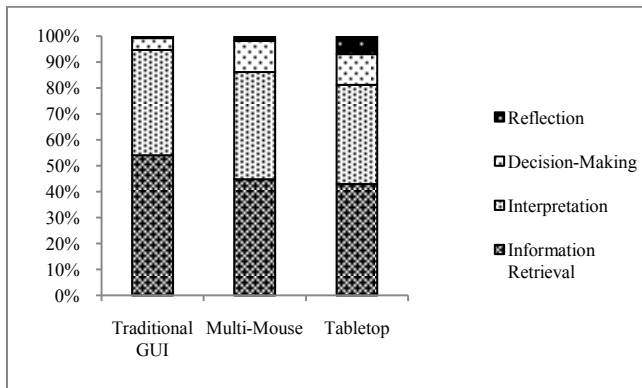


Figure 4, Time per activity relative to total session time.

Procedure

A between-subjects study design was used where 8 dyads of students were randomly assigned to one of the three experimental conditions. Sessions were held in a quiet laboratory space. Participants were given a 10 minutes tutorial demonstrating how to use the system. Then they received instructions to the first subtask. After completing the first subtask, participants received instructions to the second subtask. Upon completion of each subtask participants filled out a post-task questionnaire. The experiment concluded with a short debriefing. Each session lasted around 90 minutes. All sessions were videotaped.

Mental effort

Subjective mental effort data was collected from users following each sub-task using a twenty-step bipolar scale [26]. Participants gave a score from 0 to 100 (assigned to the nearest point 5) for stress and workload. Figure 5 shows the results. A one-way ANOVA analysis found that there were statistically significant differences between the experimental conditions for *stress* (task 1 $f(2,45)=8.573$ $p=0.001$, task 2 $f(2,45)=8.572$, $p=0.001$) and *workload* (task 1 $f(2,45)=3.283$ $p<0.05$). In task 2, differences in workload were weakly significant ($p<0.1$). Post-hoc pair-wise comparisons (using Bonferroni-corrected t-tests) showed that in task 1 the traditional GUI system resulted in significantly more stress than both the tabletop and multi-mouse GUI systems ($p=0.0004$ and $p=0.0020$, respectively) and that in task 2 the traditional GUI was more stressful

than the multi-mouse system ($p=0.0017$). It also found that in task 1 the traditional GUI system resulted in significantly greater workload than the tabletop ($p=0.0117$). In the words of one traditional GUI user: “Those are like the least-effective programs of all times. I remember using them in Bio and being like ‘I’m going to kill someone’.”

Time-on-Task

Time-on-task data was collected by an observer. Table 3 shows the average time participants spent on each sub-task. There is no statistically significant difference in the total amount of time taken to complete the sub-tasks between the different conditions. However, there were differences in how much time participants spent on each activity type depending on condition. Figure 4 shows breakdown for each type of activity relative to overall session time. A one-way ANOVA analysis shows that the average time spent on *information retrieval* and on *reflection* was significantly different between the experimental conditions ($f(2,21)=6.06$ $p<0.05$ and $f(2,21)=18.97$ $p<0.01$, respectively). Post-hoc pair-wise comparisons (using Bonferroni-corrected t-tests) showed that more time was spent on information retrieval in the traditional GUI than in the tabletop condition ($p=0.004$) and that more time was spent on reflection in the tabletop condition than in the traditional GUI and in the multi-mouse GUI ($p=0.0001$, $p=0.0008$ respectively).

	Traditional GUI M (SD)	Multi-mouse GUI M (SD)	Tabletop M (SD)
Score	66.37 (10.8)	83.87 (9.13)	85.75 (11.8)
Time			
Task 1	27:24 (9:35)	27:58 (10:55)	21:25 (2:41)
Task 2	36:13 (10:25)	28:15 (11:01)	27:38 (5:10)
Participation			
Verbal	6.63 (2.2)	6.25 (2.0)	7.25 (1.8)
Physical	2.38 (0.9)	4.75 (2.2)	14.9 (7.5)
Touch rate	1.2 (0.56)	3.4 (1.96)	13.3 (7.3)
Equity			
Verbal	0.05 (0.04)	0.1 (0.1)	0.07 (0.1)
Physical	0.14 (0.05)	0.12 (0.1)	0.15 (0.06)
Touch	0.27 (0.12)	0.15 (0.1)	0.17 (0.06)

Table 3, Descriptive statistics of experiment results.

Obtainment of Learning Goals

We assigned a score (on a scale of 0 to 100) to each dyad’s biology questionnaire based on the correctness of their answers. Table 3 shows the results. Pair-wise comparisons (using Bonferroni-corrected t-tests) showed that the scores for the two G-nome Surfer conditions (multi-mouse GUI and tabletop) were significantly higher than those of the traditional GUI ($p=0.0035$ and $p=0.0041$, respectively).

Attitude

Subjective *attitude* data was collected from users following each sub-task using a twenty-step bipolar scale. Participants gave a score from 0 to 100 for confidence and enjoyment. Figure 5 shows the results. A one-way ANOVA analysis

found statistically significant differences between the conditions for enjoyment (task 1 $f(2,45)=11.011$, $p<0.001$, task 2 ($f(2,45)=3.216$, $p<0.05$)). Differences in confidence levels found to be weakly significant for both subtasks ($p<0.1$). Post-hoc pair-wise comparisons (using Bonferroni-corrected t-tests) showed that the two G-nome Surfer conditions resulted in significantly more enjoyment than the traditional GUI ($p=0.0015$ and $p=0.0006$, respectively). In the words of one tabletop user: “Working on the big canvas is much more fun than on the computer!”

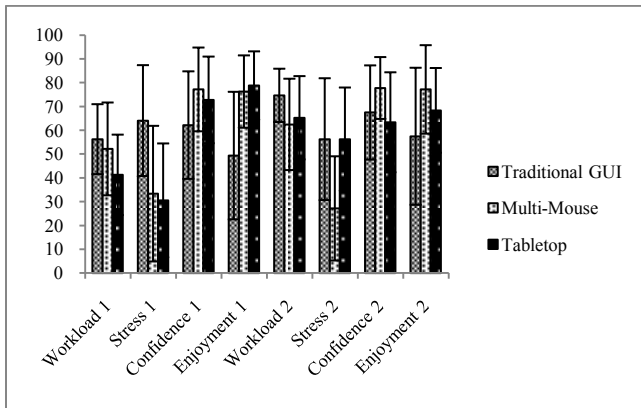


Figure 5, Subjective workload and attitude data.

Participation

Similar to Harris et al. [7], we calculated physical and verbal levels of participation for each session as proportional measures: the number of gestures and touch events per minute, and the number of utterances per minute respectively. Table 3 shows the results. Inter-coder reliability based on 37.5 percent of the data was excellent with correlation of 0.98 between coders. One-way ANOVA analysis found that there are statistically significant differences in levels of physical participation between the conditions ($f(2,21)=16.9$ $p<0.05$) and in levels of touches per minute ($f(2,21)=17.3$ $p<0.01$). Levels of verbal participation and levels of ‘off-line’ gestures did not differ significantly. Post-hoc pair-wise comparisons (using Bonferroni-corrected t-tests) showed that the tabletop condition resulted in statistically significant higher levels of physical participation as well as higher touch rates than both the traditional GUI and the multi-mouse GUI conditions ($p<0.01$ for all cases). While the significantly higher touch rates in the tabletop were expected compared to the traditional single-touch GUI condition, the comparison with the multi-mouse GUI suggests that direct touch combined with a horizontal display promotes touch. We observed that in the tabletop condition participants manipulated information artifacts – moving, resizing, and rotating - to a greater extent than in the other conditions.

To quantify *equity of participation* we calculated three Gini Coefficients [7] for each session - for verbal participation, for physical participation, and for touch rates. Table 3 shows the results. A one-way ANOVA analysis found no significant differences in equity of verbal or physical

participation between the experimental conditions. A nearly significant effect ($p=0.054$) was found for equity of touch rates: both G-nome Surfer conditions were more equitable than the traditional-GUI condition. Given that these conditions support simultaneous input this result is not surprising. The lack of significant differences in the equality of physical and verbal participation between the conditions suggests that when working collaboratively with a single mouse, users that do not control the mouse participate actively through gestures and talk.

Nature of Discussion

Figure 6 shows the proportional distribution of talk categories across experimental conditions. One-way ANOVA analysis found significant effects for the amount of utterances in the following categories: *insight* ($f(2,21)=10.2$ $p<0.01$), *coordination* ($f(2,21)=24.95$ $p<0.01$), *syntax* ($f(2,21)=13.7$ $p<0.01$), and *disengagement* ($f(2,21)=7.95$ $p<0.01$). Inter-coder reliability based on 37.5 percent of the data was excellent with 91 percent agreement. Pair-wise comparisons (using Bonferroni-corrected t-tests) show that the number of insights in the tabletop condition is significantly higher than both in the traditional GUI and the multi-mouse GUI ($p=0.002$ and $p=0.003$ respectively). The comparisons also show significantly more coordination talk in the tabletop condition than in both the traditional GUI and the multi-mouse GUI ($p=0.001$ and $p=0.005$ respectively) and significantly more coordination talk on the multi-mouse GUI than on the traditional GUI ($p=0.007$).

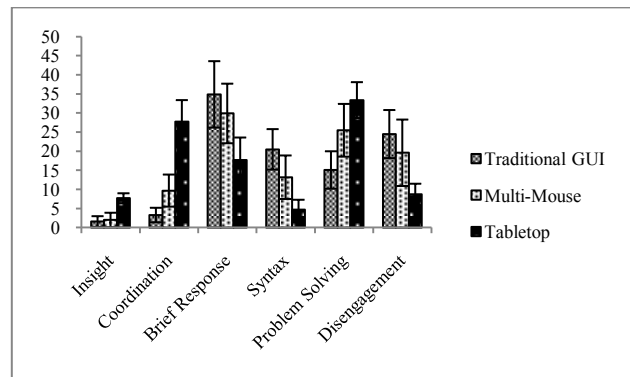


Figure 6, Distribution of talk categories per condition.

These results suggests that in the tabletop condition users tend to articulate and reflect on their insights more than in the traditional GUI and multi-mouse conditions where users often quietly write an answer and move on, keeping discussion brief. These findings also indicate that in the tabletop condition, more so than in other conditions, users establish task division through continuous discussion. We observed that in this condition participants define and take on individual subtasks as they go.

Moreover, we found that the amount of syntax utterances in the tabletop condition is significantly lower than the traditional GUI ($p<0.001$) and nearly significantly lower

than the multi-mouse condition ($p=0.020$). While a difference between the two G-nome Surfer conditions and the traditional GUI is expected, the difference between the multi-mouse and tabletop condition suggests that users found direct-touch more intuitive to use than multiple mice. We observed that in the multi-mouse condition participants used both talk and gestures in an attempt to decode which user controls which cursor. Finally, we found that the disengagement level in the tabletop condition was significantly lower than in the traditional GUI ($p=0.003$) and weakly lower than multi-mouse conditions ($p=0.03$).

Profile	Verbal indicators	Physical indicators
Turn takers: GUI: 25% Multi-mouse: 37.5% Tabletop: 62.5%	Both users make and accept suggestions and observations; Users are respectful; Decisions are reached through critical discussion.	Users share control of the mouse (traditional), coordinate multi-touch gestures and simultaneously manipulate information artifacts (tabletop, some multi-mouse). Referential gestures are used to ground talk.
Driver-Navigator: GUI: 37.5% Multi-mouse: 25% Tabletop: 37.5%	Both users are engaged: the navigator contributes with suggestions and observations; the driver listens and coordinates with navigator before performing actions.	The driver performs most actions; the navigator uses referential gestures, pointing, and continuous touch (on tabletop) to ground talk.
Driver-Passenger: GUI: 25 % Multi-mouse: 25% Tabletop: 0%	The driver is fully engaged; the passenger is not focused on the task. The driver does not explain his/her actions and often ignores or dismisses suggestions.	The driver performs most actions. The passenger rarely participates; at times the driver uses physical blocking and undoing.
Independent: GUI: 12.5 % Multi-mouse: 12.5 % Tabletop: 0%	User are absorbed in their own activity; minimal verbal communication.	Users observe each other's actions and sometimes copy them. Control over the mouse is switched after each activity or subtask.

Table 4, Collaboration profiles.

Collaboration

To better understand the *process of collaboration* we analyzed observational notes and video recordings. In our analysis, we considered the process dimensions introduced by Meier et al [16] and by the CLM framework [3] to identify verbal and physical indicators associated with effective collaborative learning. Based on these indicators as well as on quantitative participation measures, we identified four collaboration profiles that transcend interaction styles but are expressed with some differences across the experimental conditions. These include turn takers, driver-navigator, driver-passenger and independent. Table 4 describes verbal and physical indicators for each collaboration profile and specifies the percentage of each

profile per condition. Inter-coder reliability based on 37.5% of the data was good ($Kappa=0.65$).

Problem Solving Strategies

We observed different problem solving strategies across the experimental conditions. In the tabletop condition, about three quarters of the dyads used an *accumulation and comparison* strategy, where participants first accumulated information *vertically* – gathering different pieces of information for each biological entity and then rearranged the information spatially to compare information across biological entities (i.e. horizontally). Decisions were made through an extensive critical discussion while referring to the information. The remaining quarter of the dyads on this condition used a stepwise elimination strategy.

In the multi-mouse condition, about two third of the dyads used a *stepwise elimination* strategy where participants accumulated information vertically, focusing on a single biological entity (e.g. gene) and then eliminated possibilities through discussion (while removing information artifacts from the screen) before they moved on to investigate the next entity. In some of these dyads, there has been also a brief discussion to summarize all findings at the end of the task, but the information was no longer presented on the screen at this phase. The rest of the dyads in this condition did not exhibit a set strategy. Instead, they accumulated some data, started stepwise elimination, and then backtracked to retrieve information they already eliminated. Dyads that used *backtracking* were occasionally overwhelmed. In the traditional GUI condition, about half of the dyads exhibited frequent backtracking. Quarter of the dyads used stepwise elimination, while the remaining quarter used accumulation and comparison. These dyads used browser tabs to save different pieces of information, and used paper to write down important data.

Discussion

This study takes a holistic approach to understanding the collaborative learning process mediated by the experimental conditions. We assessed the efficiency and effectiveness of collaborative learning while considering performance, verbal and physical indicators, and subjective users' data. Our findings indicate that G-nome Surfer reduces users' stress levels and workload compared to current state-of-the-art tools as well as improves students' performance and attitude. Since we did not find significant differences between the two G-nome Surfer conditions in respect to these strengths, we attribute them to G-nome Surfer's visual design, interaction techniques, and integrated workflow.

We also found that verbal and physical participation were highly equitable across experimental conditions. This suggests that multi-mouse GUI and multi-touch tabletop do not provide benefits in terms of *equality* of participation. However, quite expectedly, these conditions result in more equitable touch rates. No significant differences related to equality of participation were found between the multi-mouse GUI and the tabletop condition.

We did find the multi-mouse GUI and tabletop conditions to differ in several areas. First, we measured significantly higher levels of *physical participation* in the tabletop condition. These are attributed to high touch rates that were expressed by increased spatial manipulation. Several theories of embodied cognition suggest that spatial manipulations or “thinking with our hands” can help reasoning about abstract concepts [14]. Second, we found that the tabletop condition encourages *reflection*. This was evident by the longer time spent on reflection activities compared to the other conditions, the significantly higher number of articulated insights, and the problem solving strategy of accumulation and comparison that was used by the majority of tabletop dyads. Since research indicates that student’s understanding of the nature of science is enhanced through reflection [30] this is an important strength. Third, we found that the tabletop condition fosters a more *effective collaboration* where both participants are actively engaged while maintaining joint focus on the task. This is evident by the turn-taking collaboration style exhibited by the majority of dyads in this condition, the significantly higher number of coordination utterances, and the significantly lower number of disengagement utterances. Effective collaboration is indicative of effective collaborative learning [13]. Finally, the significantly lower number of syntax related utterances and the shorter time spent on information finding suggests that a multi-touch tabletop is easier and more natural to use.

These findings support our hypothesis that the tabletop condition benefits collaborative learning by facilitating a more *effective* collaborative learning process. However, while some findings indicate that multi-touch tabletop interface is more intuitive to use, the study does not provide conclusive evidence that the tabletop condition facilitates a more *efficient* learning process than a multi-mouse GUI (in terms of invested mental effort). Since adoption barriers for tabletop interfaces are relatively high due to cost, size, and to perceived lack of flexibility, highlighting the strengths of tabletop interaction compared to multi-mouse GUI have implications for choosing an interaction style for supporting college-level collaborative learning.

This study has several limitations that point towards future work. First, we studied one-time use in laboratory settings. Additional studies of longitudinal use are necessary in order to determine to what extent the findings are affected by novelty. Second, our measurement of learning outcomes based on a short-term study is limited and does not assess individual learning or participants’ ability to apply their learning. In the future we plan to use additional assessment instruments to further measure learning outcomes.

CONCLUSIONS AND FUTURE WORK

While the domain of genomics provides the frame for this work, this research makes three contributions to tabletop interaction and computer-supported collaborative learning in general. First, we identify design requirements for

supporting collaborative learning in areas that use vast amounts of heterogeneous information. Second, we demonstrate a tabletop interface, G-nome Surfer, which employs multi-touch and tangible interaction techniques to explicitly support collaborative learning in college settings. G-nome Surfer improves performance, reduces workload, and increases enjoyment. Finally, we present findings from a study that identifies four educational benefits of tabletop interaction compared to a multi-mouse GUI: 1) increasing physical participation, 2) encouraging reflection, 3) fostering effective collaboration, and 4) facilitating more intuitive interaction.

In the future, our goal is to enhance scientific discovery and education in areas that explore vast amounts of data. We intend to integrate G-nome Surfer into undergraduate biology courses, further evaluate it in a longitudinal study, and deploy it in research labs that train future scientists.

ACKNOWLEDGMENTS

We thank our collaborators Carol Ann Paul and Ginny Quinan from the Neuroscience program at Wellesley College. We also thank Andrea Sequeira, Yair Benita, and the instructors and students participated in our study. This work is partially funded by NSF grant no. IIS-1017693.

REFERENCES

1. Antle, A. N., Droumeva, M., and Ha, D. Hands on what? Comparing children’s mouse-based and tangible-based interaction. *Proc. of IDC 2009*. ACM Press (2009), 80-88.
2. Bolchini, D., Finkelstein, A., Perrone, V., Nagl, S. Better bioinformatics through usability analysis, *Bioinformatics*, 25(3) (2009), 406-12.
3. Fleck, R., Rogers, Y., Yuill, N., Marshall, P., Carr, A., Rick, J., and Bonnett, V. Actions speak loudly with words: unpacking collaboration around the table. *Proc. of Interactive Tabletops and Surfaces*, (2009).
4. Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Bötschi, K., Voegtli, B., and Juchli, P. Tangible user interface for chemistry education: comparative evaluation and re-design. *Proc. of Human Factors in Computing Systems*, ACM Press (2007).
5. Gillet, A., Sanner, M., Stoffler, D., Goodsell, D., and Olson, A. Augmented Reality with Tangible Auto-Fabricated Models for Molecular Biology Applications. *Proc. of Visualization 2004*, IEEE (2004).
6. Gläser, T., Franke, J., Wintergerst, G., and Jagodzinski, R. Chemieraum: tangible chemistry in exhibition space. *Proc. of TEI 2009*, ACM (2009), 285-288.
7. Harris, A., Rick, J. Bonnett, V., Yuill, N., Fleck, R., Marshall, P., and Rogers, Y. Around the table: Are multiple-touch surfaces better than single-touch for children’s collaborative interactions? *Proc. of CSCW* ACM Press (2009), 335-344.

8. Hart, S.G., and Stavenland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in *Human Mental Workload*, Elsevier (1988), 139-183.
9. Horn, M.S., Tobiasz, M., Shen, C. Visualizing Biodiversity with Voronoi Treemaps. In Proc. Sixth International Symposium on Voronoi Diagrams in Science and Engineering (2009).
10. Hornecker, E., Marshall, P., Dalton, N.S., Rogers, Y. Collaboration and Interference: Awareness with Mice or Touch Input, *Proc. of CSCW*, ACM Press (2008).
11. Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. Reality-based interaction: a framework for post-WIMP interfaces. Proc. of *Human Factors in Computing Systems*, ACM Press (2008).
12. Kirk, D., A. Sellen, S. Taylor, N. Villar, and S. Izadi. Putting the physical into the digital: Issues in designing hybrid interactive surfaces. *Proc. of HCI*, ACM Press (2009), 35-44.
13. Kirschner, F., Paas, F., and Kirschner, P.A. A cognitive load approach to collaborative learning: United brains from complex learning. *Educational Psychology Review*, (2009) 21, 31-42.
14. Klemmer, S. R., Hartmann, B., and Takayama, L. How bodies matter: five themes for interaction design. *Proc. of DIS 2006*, ACM Press (2006).
15. Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. Reality-based interaction: a framework for post-WIMP interfaces. *Proc. Human Factors in Computing Systems* (2008).
16. Meier, A., Spada, H., & Rummel, N. A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning* (2007), 2, 63-86.
17. Mirel, B. Supporting cognition in systems biology analysis: findings on users' processes and design implications, *J Biomed Discov Collab* (2005), 27(2).
18. Morris, M.R., Brush, A.J.B., and Meyers, B. Reading revisited: evaluating the usability of digital display Surfaces for Active Reading Tasks. *Proc. of IEEE Tabletop* (2007), 79-86.
19. Ochs, E., Jacoby, S., and Gonzales, P. Interpretive journeys: how physicists talk and travel through graphic space. *Configurations*, (1994) 2, 151-171.
20. Patten, J., and Ishii, H. A comparison of spatial organization strategies in graphical and tangible user interfaces, *Proc. of Designing Augmented Reality Environments* (2000), 41-50.
21. Paul, C.A. and Quinan, G. Behavior, neuroanatomy, genomics: what can we learn from mouse mutants?, *Proc. of Workshop on Teaching Big Sciences at Small Colleges*, (2009).
22. Piper, A.M., Hollan, J. Tabletop displays for small group study: affordances of paper and digital materials. *Proc. of CHI 2009*, ACM Press (2009).
23. Rogers, Y. and Lindley, S. Collaborating around large interactive displays: which way is best to meet? *Interacting with Computers*, 2004.
24. Roth, W. M. Gestures: their role in teaching and learning. *Review of Educational Research*, (2001) 71, 365-392.
25. Ryall, K., Forlines, C., Shen, C., Morris, M.R. Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. *Proc. of CSCW 2004*. ACM Press (2004), 284-293.
26. Rubio, S., Diaz, E., Martin J., and Puente, J.M. Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX and workload profile methods, *Appl. Psychol. Int. Rev.* (2004) 53, 61-86.
27. Saraiya, P. and North, C. and Duca, K. An insight-based methodology for evaluating bioinformatics visualizations, *Visualization and Computer Graphics*, (2005) 11(4), 443-456.
28. Schkolne, S., Ishii, H., Schroder, P. Immersive design of DNA molecules with a tangible interface, *Visualization*, IEEE (2004), 227- 234.
29. Shaer, O., Kol, G., Strait, M., Fan, C., Grevet, C., and Elfenbein, S. G-nome surfer: a tabletop interface for collaborative exploration of genomic data. *Proc. of Human Factors in Computing Systems*. ACM Press (2010).
30. Singer, S.R., Hilton M.L., and Schweingruber, H.A. America's lab report: investigations in high school science, National Research Council (2005).
31. Tanner, K., Chatman, L. S., and Allen, D. Cooperative learning in the science classroom - beyond students working in groups. *Cell Biol. Educ.* (2003) 2,1.
32. Teaching big science at small colleges: genomics collaboration. <http://serc.carleton.edu/genomics/>
33. Woolgar, S. Time and documents in researcher interaction: some ways of making out what is happening in experimental science. *Human Studies* 10, 2-3(1988).