From Big Data to Insights: Opportunities and Challenges for TEI in Genomics

Orit Shaer[1], Ali Mazalek[2], Brygg Ullmer[3,4], Miriam Konkel[5]

Dept. of Computer Science [1] Wellesley College Wellesley, MA, USA

Synaesthetic Media Lab[2] Digital Media Georgia Institute of Technology Atlanta, GA, USA

School of EECS[3], Center for Computation and Technology[4], and Dept. of Biological Sciences[5] Louisiana State University Baton Rouge, LA, USA

oshaer@wellesley.edu, mazalek@gatech.edu, ullmer@lsu.edu, konkel@lsu.edu

ABSTRACT

The combination of advanced genomic technologies and computational tools enables researchers to conduct large-scale experiments that answer biological questions in unprecedented ways. However, interaction tools in this area currently remain immature. We propose that tangible, embedded, and embodied interaction (TEI) offers unique opportunities for enhancing discovery and learning in genomics. Also, designing for problems in genomics can help move forward the theory and practice of TEI. We present challenges and key questions for TEI research in genomics, lessons learned from three case studies, and potential areas of focus for TEI research and design.

Author Keywords

Tangible interaction, interactive tabletops, computational genomics, biology, big data, scientific discovery, learning.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces---*input devices and strategies, interaction styles*; J.3 [Computer Applications]: Life and Medical Sciences---*biology and genetics*.

General Terms

Design, Human Factors.

INTRODUCTION

Advances in genomic technologies have transformed biological inquiry and have the potential to alter medical practice to offer much-improved health care [7]. Genomic and biological technologies are also positioned to address some of the most pressing problems of our times, including food and clean water shortage, as well as increased demand for alternative energy sources [6].

The application of genomic technologies has opened new interfaces between biology and computer science, fueling fields such as bioinformatics that enable biological questions to be tackled computationally [7]. Resonant with broader evolutions in science, the study of genomes is evolving on an equal footing in theory, experimentation, and computation. The combination of advanced genomic technologies (e.g. second generation DNA sequencing) and powerful computational tools has facilitated biological investigations in a manner and scale previously not possible [24]. Rather than only small-scale analyses, researchers now often conduct large-scale experiments in which information from multiple genes and genomes is measured, recorded, analyzed, and stored in databases.

The bottlenecks and challenges along the path to transforming the "big data" generated by these experiments into biological insights – including observations about the data that constitute units of discovery [30] – are formidable and numerous [7]. For one, data analysis is now replacing data generation as the rate-limiting step in genomic research [24]. Large-scale genome research efforts, which involve generating and interpreting data at an unprecedented scale, have brought into focus the need for new computational tools that facilitate meaningful analyses.

The deluge of information generated by emerging genomic technologies has driven a change not only in scale of investigations, but also in the tools used by biologists. Next to the pipette and pen, a web browser is today one of the most widespread biology tools, as it provides access to powerful computational resources [35]. Web technologies have been massively adopted toward facilitating improved access for biologists, who presently are often not computer experts. However, today's computational tools show severe limitations in persistence, usability, and support for collaboration and high-level reasoning [3, 21, 45].

Based on our experiences working at the intersection of Genomics and TEI both in the capacity of interaction designers [35-37, 43, 49] and genomic investigators [11, 16, 27], we propose that the research area of TEI offers special opportunities for enhancing discovery and learning in genomics. We see these prospects arising from design and realization of interfaces for exploration, interpretation, organization, manipulation, and sharing of vast dataspaces mediating the construction of new insights. We further suggest needs for innovative interfaces enhancing discovery

TEI 2013, Feb 10-13, 2013, Barcelona, Spain.

processes involving big data present an opportunity to drive forward the field of TEI. Participants of the inaugural TEI conference panel, in discussing research agendas for the field, suggested: "more complex computation should be occurring behind the tangible interface, instead of only one-to-one input-output." [13]. We view developments in computational sciences and TEI as intertwined, with both Computational STEAM (science, technology, engineering, art, and math) and TEI prospectively advanced by leveraging old and new synergisms between scientific disciplines, arts, and culture [43].

In this paper we discuss opportunities and challenges for the application of TEI research to computational genomics. Our contribution in this work is threefold. First, we analyze and characterize the genomics problem domain, deriving challenges and key questions for TEI research. Second, we present lessons from three case studies at the intersection of TEI and computational genomics/biology. Finally, drawing on our described experiences, we discuss areas for design focus that show potential to advance discovery and practices in genomics, as well as to contribute new knowledge to the theory and methods of TEI. We begin with a brief survey of related work in TEI for the sciences.

RELATED WORK

A number of systems illustrate possibilities for supporting scientific discovery and higher education with TEI. Brooks et al. [4] developed an early haptic display for scientific visualization. Gillet et al. [9] presented a tangible user interface for molecular biology that used augmented reality technology to view 3D molecular models. Schkolne et al. [31] developed an immersive tangible interface for the design of DNA molecules. While these systems highlight potential benefits of TEI for scientists, they focus on the representation of objects with inherent physical structure. We are interested in a broader use case, where abstract information is represented and manipulated.

Several projects investigate augmented capture and situated access to biological data. Labscape [2] is a smart environment for cell biology labs. ButterflyNet [50] is a mobile capture and access system for field biologists. Mackay et al. [17] and Tabard et al. [39] explore the integration of biologists' notebooks with physical and digital information sources. Our efforts have been oriented toward the (computational) workbench, but are synergistic with past and future paper-entangled opportunities.

To date, several systems have been developed to facilitate collaboration among scientists across large displays and multi-touch tables. WeSpace [46] integrates a large data wall with a multi-touch table and personal laptops. TeamTag [29] allows biodiversity researchers to collaboratively search, label, and browse digital photos. eLabBench [40] investigated tabletop interfaces as interactive wet lab benches. Kuznetsov et al. explored the development of artifacts for supporting DIYbio [15].

Finally, TEI systems have also illustrated the potential to support science education. We discuss several relevant to genomics. Augmented Chemistry [8] is a tangible user interface for chemistry education. Involv [12] is a tabletop interface for exploring the Encyclopedia of Life that shares our challenge of creating effective interaction techniques for large data spaces. PhyloGenie [32] is a tabletop interface for collaborative learning of phylogeny through guided activity. We are interested in the development of interfaces that promote open-ended hands-on inquiry.

CHALLENGES FOR TEI

The field of genomics encompasses a broad scope of scientific inquiry focusing on the organization and function of genomes. This wide-ranging scientific discipline gives rise to a set of design challenges for TEI.

Challenges #1: Diverse Audience

At least four distinctive user groups are involved in the domain of genomics. Each is marked by unique needs, and thus likely best serviced by distinct design criteria, which in turn may be realized in different tools:

First, *genomic scientists* are domain experts whose goal is to derive scientific insights from large-scale data sets. Genomic scientists work on a broad variety of tasks, each with its own biological entities of interest and analytic workflows. Three core tasks are typically included in a broad range of investigations [24, 35]: 1) Annotating sequence data; 2) Browsing annotations mapped to a reference genome; and 3) Comparing genomic sequences. Scientists require tools that combine powerful automated computational analysis with support for manual inspection, interpretation, and high-level reasoning.

Second, *future scientists* are students early in their scientific career. Training in genomics often includes research experiences that require access to and the manipulation of large-scale data sets through the use of sophisticated computational methods and interfaces. However, current interfaces pose a high threshold for novice users and do not support important aspects of learning such as inquiry-based high-level reasoning, the development of process knowledge, and collaborative learning [21, 25, 35].

Third, *citizen scientists* are people with little or no formal scientific training who seek to make meaningful contributions to research. While some crowd-sourcing projects provide novel interfaces for citizen scientists to help analyze data collected by scientists (e.g. [5]), there is a growing movement to empower people to gather, control, and analyze their own data. Prompted by the increasing availability of consumer DNA tests, citizen science research in genomics is often motivated by questions people have about their health risks and how to prevent them. For example, DIYGenomics is a citizen science organization that facilitates sharing genetic information, tracking data, and forming collaborations. The work of citizen scientists in genomics sometimes leads to scientific contribution (e.g. [18]). Still, existing platforms only serve as starting points

in addressing the growing need for tools that will make it easier for people to contribute to research.

Fourth, the *general public* refers to the need to increase the awareness of citizens and policy makers to the opportunities and challenges of genomic research and its potential to transform public health. For this audience, the challenge is to develop culturally competent educational tools and material that communicate complex data to non-scientists.

This diverse audience opens new opportunities and questions for TEI research. For one, while TEI researchers have focused substantial effort on novice users and "walk-up-and-interact" systems, little work investigates the design of TEI systems for expert users. What is the potential of TEI systems that involve experts interacting with big data? What sort of custom hardware, visual design, and user training would they require? Also, while several studies have investigated the effects of TEI on learning, most have focused on children. How can interactive surfaces be used to help students in higher education learn complex concepts? Finally, how can artistic and cultural artifacts be used to communicate scientific data in outreach programs?

Challenges #2: Scale

Current genomic research combines unprecedented capacity for data generation with powerful computational tools, resulting in large-scale investigations. We discuss the scale of the data and the collaboration such investigations entail.

Data: The recent introduction of next-generation sequencing technology and the rapidly falling costs of DNA sequencing (far faster than Moore's Law [19]) are rapidly changing the landscape of genomics. Many genomic research studies now involve the detailed study of multiple (sometimes thousands) whole genomes. A single genome often contains several billion DNA base pairs. An assembled genome could be seen as a linear sequence of genomic letters for each chromosome. In practice, reality is often more complex. With next-generation sequencing, a genome is fragmented into millions of short DNA fragments. A second-generation DNA sequencer identifies (with some errors) the sequence of base pairs on each of these DNA fragments. The sequence reads of the most commonly used sequencing technologies are relatively short. These can be used for numerous approaches and analyses – e.g., through genome assembly, or detecting single nucleotide variations, repetitive (non-gene) content, etc.

TEI research has not yet addressed scale and complexity at these magnitudes. Challenges include development of interfaces going beyond one-to-one mapping, providing means for searching, comparing, and sharing big data. What representations are appropriate for large, abstract data? What interaction techniques could potentially reduce the mental workload associated with handling big data?

Collaboration: Research in genomics relies increasingly on large-scale projects with hundreds of collaborators distributed across dozens of institutions spanning the globe. For example, the International Cancer Genome Consortium (ICGC) [7] leads a long-term effort encompassing over 20 projects from 14 countries, with hundreds of collaborators. Within these large collaborative efforts, smaller co-located often-interdisciplinary teams carry out individual projects. Collaborative work in such teams is typically based on emails, weekly conference calls, and face-to-face meetings [35]. Researchers store their results in a shared database. Communication is often the key to project success.

Despite the importance of collaboration, current bio-informatics tools do not support collaborative exploration. While TEI has supported small co-located collaboration over short time scales, a challenge includes how to manage the work across entire teams, given their global nature and large temporal scale. For example, how to provide activity and progress visualizations across separate systems? How to increase awareness across time and location?

Challenges #3: Heterogeneous Data

Biologists combine multiple forms of evidence to discover connections and causal relationships, as well as to examine information at different levels of granularity [21, 35]. For example, the investigation of cancer genomes includes studying local changes such as substitution mutations and insertions/deletions, as well as global changes such as chromosomal rearrangement and nucleic acids of foreign origin (e.g. oncogenic viruses) [7]. This requires users to move back and forth between levels of granularity without losing context – from viewing an entire genome or a large chromosome area to the base-pair level and back. Considering other data, such as alterations in the mechanisms that regulate gene expression, is also important in the investigation of cancer tumors, but requires the use of different bioinformatics tools. In addition, biologists often compare the genomes of multiple tumors. To do so they load multiple genomes to a single genome browser window, which results in displaying a large (and often overwhelming) number of parallel tracks for a short stretch of DNA.

The need to link various datasets and tools to gain insight into complex systems, poses several challenges for TEI. For example, how to design interfaces that facilitate the manipulation of big data in a way that highlights connections between multiple forms of evidence? Or, how to represent abstract heterogeneous information while at the same time promoting learning and discovery?

CASE STUDIES

Following, we describe three case studies focused on computational genomics, which attempt to tackle the challenges discussed above. While these systems have limitations, they highlight the potential of leveraging TEI techniques to facilitate effective interaction with big data.

CS1: Tabletop genome browsing & primer design

G-nome Surfer Pro [37] is a tabletop interface for browsing *prokaryotic* genomic data. It was designed to support teams of students participating in authentic scientific inquiries. Its design was motivated by the lack of bioinformatics tools,

which support integrated workflow while facilitating collaboration, learning, and high-level reasoning [3, 21].

Prokaryotic genomes differ from eukaryotic genomes. They are smaller and typically contain a single gene-rich circular piece of chromosomal DNA. Considering these differences between eukaryota and prokaryota, the development of G-nome Surfer Pro, while drawing on earlier versions of G-nome Surfer [35, 36], required the design of new visualizations and interaction techniques. The design of G-nome Surfer Pro (see Figure 1) is a result of a participatory design process where we partnered with domain experts. Our goals included: 1) Lowering the threshold for using advanced bioinformatics tools; 2) Supporting an integrated and flexible workflow; and 3) Fostering collaboration and reflection. Our design was informed by existing research, indicating that tabletop interfaces support collaboration through visibility of actions and egalitarian input [14, 20] and afford distributed cognition [26].

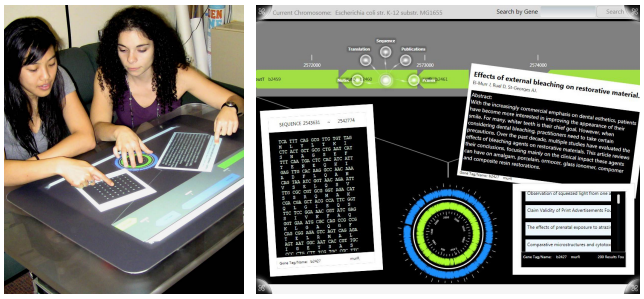

Figure 1. Students using G-nome Surfer Pro (left): displaying chromosome visualizations, DNA sequence, and related publications (right).

The current prototype of G-nome Surfer Pro utilizes multi-touch interaction techniques with a visual genomic map: a circular genome visualization (i.e. wheel) displaying an overview of the entire genome, along with a magnified view of a slice of the chromosome. Users are able to pan the chromosome left and right either by rotating the chromosome wheel (for coarse navigation) or by using a flick gesture on top of the magnified slice (for fine navigation). A visual indicator links the wheel and the slice, helping users to maintain a sense of location. Users can retrieve genomic sequences, access GenBank notes, search for publications, or access a primer designer. Users can spatially manipulate, annotate, and compare information artifacts. Figure 1 shows a screen capture from G-nome Surfer Pro that displays a genomic map, an aligned sequence, and related information artifacts.

In addition to facilitating genome browsing, G-nome Surfer Pro also introduces support for manipulating DNA through the task of primer design. This task involves the identification and testing of short sequences of DNA marking the start and end of a particular region of DNA sequence. G-nome Surfer Pro allows scientists to test and manipulate their primers through direct manipulation, giving the user control over their primer design process.

G-nome Surfer Pro was evaluated with 14 undergraduate research students. Findings indicated that G-nome Surfer Pro was an effective tool for complex interaction with large amounts of *prokaryotic* genomic data. In particular, we found that the multi-touch interaction seemed natural to users while the horizontal surface facilitated effective collaboration styles. The evaluation of G-nome Surfer 2.0, a sister application which was designed for the exploration of *eukaryotic* genomics [36, 37], also highlighted how users applied spatial problem solving techniques facilitated by the large horizontal surface to explore large data sets. Such techniques include accumulation and arrangement, as well as side-by-side comparison, of artifacts.

During the two years in which we developed different prototypes of G-nome Surfer [35, 36, 37], we experimented with various visual and tangible representations of different aspects of the genome browsing workflow. For example, we used tangible test tubes as containers for particular data sets [35]. We also used iconic and playful tangibles for invoking certain computational functions. While we demonstrated that tangible interactions can facilitate immediate, visible, and easily reversible manipulations, it also became evident that further investigations are required for identifying combinations of control (multi-touch and tangible) and representations (physical and visual) that allow for open-ended, powerful, and expressive inquiries.

CS2: Tangibles-targeted computational genomics

Our second case study considers computational genomics work conducted with a vision toward tangible interaction support. We briefly describe two efforts, each a part of large international genome consortia.

Our first major experience in a computational genomics effort engaged the rhesus macaque genome, together with ~250 collaborating scientists [11, 27]. We compared the rhesus macaque genome with the human and chimpanzee genome; characterized similarities and differences in a particular class of genomic content; and experimentally validated these computational results. Roughly 400,000 differences in genomic content relative to the human and chimpanzee genomes were detected and analyzed. To realize this, ~200 compute cores were used for 2-3 weeks non-stop. This is comparable to a single-processor computer running nonstop for a decade. We anticipated code parallelization would be complex; but for our highly repetitive analysis, it was relatively straightforward.

Instead, the scale and complexity of data analysis was most remarkable. For compatibility with mainstream tools, our analyses yielded several million small files (e.g., 10-15 files per gap, across 400k gaps). For flavor, even when used on a high-performance networked file system, a recursive file listing took four days to complete. In a second iteration, we used constellations of several hundred ZIP files. This aided computational efficiency, but not human navigation.

Our second major genomics experience was with the orangutan genome [16]. With the rhesus genome project, our efforts largely confirmed an anticipated result. With

orangutan, the opposite held true: our efforts yielded a highly unexpected result. Specifically, one genomic element of interest had far fewer species-specific insertions than expected. Initially, this struck us as highly unlikely, leading us to distrust our analyses. Ultimately, a world expert confirmed this result, leading to a major finding. This process increased our interest in approaches for visualizing and interactively manipulating analyses to gain confidence in unexpected findings.

Another experience related to databases. Chastened by our earlier file system experiences and inspired by the scientific database discussions of Gray et al. [10], we attempted to employ SQL databases to manage our analyses. We were working with second generation DNA sequencer datasets, each containing more than 10 million small DNA fragments. With multiple records per fragment, we faced databases with hundreds of billions of records *per dataset.*

At these scales, SQL behaviors that are desirable for more traditional datasets became problematic. For example, in order to guarantee that a database will not be corrupted by an incomplete operation, SQL databases typically guarantee provisions for "rollback". However, for our data, if one wishes to toggle a single status bit-flag across all fragments, it could take days or even weeks to execute. The SQL tool did not have profiling support. Consequently, a week of processing would pass, and it was indeterminable whether processing was nearly complete, or barely begun.

These and other experiences left us with hopes and opportunities for applying tangible interfaces to genomics. Tangible representations of multi-stage workflows could be applied to launching and tracking the progress of complex analyses, and facilitating discussions of assumptions, way-points, and results. Parameterized, annotatable visualizations of complex, large-scale datasets would be highly valuable. Weekly conference calls of genome consortia often involve 50 or more participants spanning the globe. Tools for allowing these teams to hold constructive discussions, actively manipulating assumptions, and seeing immediate consequences could have high impact.

CS3: Tangibles for visualizing systems biology

Pathways [49] is designed to support computational modeling of biochemical systems to simulate and understand phenomena such as different diseases or plant systems. The target users are researchers with engineering or mathematics backgrounds who currently model in several overlapping iterative stages. A model is constructed as a system of ordinary differential equations (ODEs) that represent reactions in the biochemical system. Experimental data is gathered from literature or experimental biology collaborators. The ODE system is solved numerically and the results are plotted as graphs, which are compared with the experimental data. The graphs rarely match at first try, and the modelers adjust parameters (e.g. concentrations of molecules) to "fit" the ODE output to the experimental data. The model must fit the global pattern of the data. For a network of N equations, this is like comparing two N-dimensional structures over time. Much of the fitting process is thus done using optimization algorithms, which try out values in the solution space with the goal of reducing the difference between the ODE and experimental curves for all equations in the network. Finally, the model is run through diagnostic tests. If these fail, the parameters need to be tuned or the pathway changed.

The modeling process involves many separate tasks: e.g. sketching, developing algorithms, analyzing output, and gathering data. A big challenge is the difficulty of finding rich and dependable data. The data used is typically very sparse. A related challenge is that a range of parameter values can generate model results that fit the data. In other words, the solution may not be unique, and so achieving fit does not necessarily mean that the model has successfully captured the biological mechanism being modeled. There is also the possibility that the network may be wrong and the modeler may need to add or remove components.

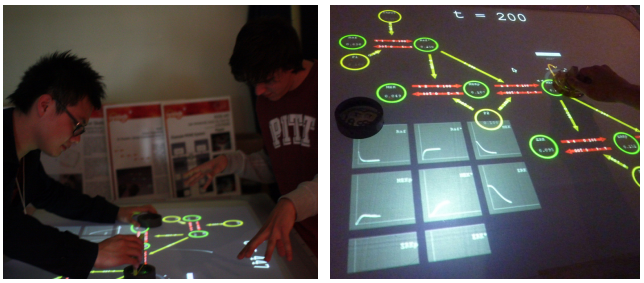

Figure 2. Two researchers using Pathways (left); a tangible dial is used to adjust the concentration of a molecule in the reaction network (right).

Pathways [49] (see Figure 2) aims to support the modeling process by providing: (1) a comprehensive representation that brings together the different stages of the modeling process, and (2) kinesthetic interaction with the system that can give the modeler an embodied sense of fit. In other words, the physical interactions with the interface should relate to the properties of the biological process being modeled, so that as the modeler gets a physical feel for the way the model works, they simultaneously develop a clearer conceptual understanding of the mechanism it represents. This design goal builds on theories of embodied cognition, which hold that sensory and motor processes play significant roles in cognitive processes [1, 47].

The first version of Pathways shows a dynamic display of the reaction network and associated graphs. More recent versions also provide a representation of the experimental data. The user manipulates the model with tangible controls and tries to bring it closer to the experimental data. Pathways still has limitations. The biggest is that the tangible manipulations currently happen in a localized manner, making it difficult to get a global embodied feel for the model. However, it has become evident that a tool like Pathways could simplify the modeling process and could also enhance collaboration between engineers and

biologists by providing them with a shared representation for thinking about the biological systems they study.

POTENTIAL AREAS FOR TEI ENGAGEMENT

Designing TEI systems to tackle problems in genomics goes beyond the application of existing approaches. Addressing the scientific challenges discussed above will necessarily challenge some of the interaction methods used in TEI systems to date. We believe that forging collaborations between TEI and genomics researchers can create a positive feedback loop that will help to move both fields forward. Drawing on our case studies, we discuss areas for design focus that can capitalize on the unique affordances of TEI. We believe that developing novel TEI designs in these areas can potentially advance discovery and practices in genomics, as well as contribute new knowledge to the theory and methods of TEI.

Area #1: Understanding complex problems

All three case studies illustrate complex problems involving biological elements and mechanisms that are still under investigation. CS3 illustrates that current computational modeling practices are difficult to follow and understand. To support understanding and discovery, tools need to provide access to relevant aspects of the problem (data, reactions, etc.) using a comprehensible representation, as well as to enable manipulation of biological data in ways that help users develop accurate mental models. TEI research has demonstrated different ways to facilitate problem solving and understanding through bodily actions, physical manipulation, and tangible representations [34]. Some of the core questions in TEI surround representation. What should be physical, and what should be digital? How should abstract concepts be effectively embodied?

Recent theories in the cognitive sciences, such as embodied cognition, provide data that illustrates the importance of the body, external artifacts, and the environment in reasoning and learning [1, 47]. In the sciences this is hardly surprising, since external artifacts, including sketches, diagrams, and physical models have long been used to support reasoning [22, 23]. TEI approaches, which employ kinesthetic, direct, gestural and spatial interactions with both physical and digital representations, are thus well suited to support scientific discovery. In TEI systems, users manipulate physical and digital artifacts with their hands, such as the data and models in CS1 and CS3. They can also arrange and manipulate the data spatially (CS1 and CS3).

Major challenges include identifying ways too combine physical and digital representations that support powerful and expressive inquiry as well as designing the interaction system such that its dynamic or movement properties relate to those of the studied biological system. If such relationships can be established, TEI systems could engage the connection between the hand, the eye, and the brain to support users' conceptual understanding.

Area #2: Visualizing biological data

Visualization is one of the areas in which the arts have forged a close relationship with science and technology (e.g. [48]), as art and design disciplines have a long history for turning ideas and experiences into expressive creations, realistic and abstract. Each of our case studies addresses different challenges for visualizing biological data. CS1 illustrates the need to navigate and view data at different scales. CS2 highlights the potentially enormous size of the data set. CS3 demonstrates the potential sparseness and unreliability of available data. All demonstrate a need to compare and analyze separate pieces of data in parallel.

The large surfaces, horizontal and vertical, that are a feature in many TEI systems are well suited to displaying and navigating large amounts of data. Still, designing and developing effective visualization techniques that can support these needs and support high-level understanding is a big challenge. Tabletop interfaces in particular can allow the combination of physical manipulation with the display of dynamic information at different scales. Parallel interactions, multi-handed or multi-user, can allow comparison and analysis of multiple pieces of data. With TEI, visualization of data can also be realized in different forms, not only on flat surfaces, but also physically embodied, or as combinations of graphical and physical artifacts (e.g., [44]). Opportunities and challenges for TEI include devising methods for realizing these visualizations, representing and swiftly transforming between multiple scales while maintaining context, perhaps enhanced by emerging sensing technologies and smart materials.

Paths forward seem likely to begin with small steps toward specific problems; then investigating how supporting visualization and interaction methods can be generalized. Some of these generalizations may draw from or contribute to other areas of interactive computational STEAM.

Area #3: Enabling large collaborations

The case studies illustrate different kinds of collaborative practices in computational biology and genomics, from small groups of students working together in CS1, to large teams of experts in international genome consortia in CS2. In the case of small groups working together on problem solving tasks, TEI systems provide form factors that foster collaboration and parallel interactions by multiple users, tabletop interfaces are well suited for this, as are combinations of tables, wall displays, personal devices, and tangible artifacts (see e.g., [33]).

One challenge for TEI is the development of systems that can support large collaborations like in CS2. One feature of these collaborations is that they often consist of smaller co-located teams working together as part of a large distributed team. Extending to large group collaboration is thus a question of connecting TEI systems for individual smaller teams across space and time, to create a shared platform for the larger distributed team. This approach has been explored by mixed presence groupware (MPG) systems,

which typically employ touch or pen interaction and support remote awareness through visual representations of remote users' arms (e.g., [41, 42]). However, MPG systems have rarely considered tangible interaction or distributed teams at more than two locations. A challenge for TEI researchers is thus to investigate how the benefits of tangible interaction for co-located collaboration might extend into remote-team collaboration situations, especially across more than two sites.

Area #4: Supporting diverse audiences

The case studies illustrate groups of users working in different contexts, from students in CS1, to experts in CS2, and both novices and experts in CS3. These audiences are also diverse in terms of their disciplinary backgrounds: biologists, engineers, mathematicians, computer scientists, and beyond. International genome consortia as shown in CS2 particularly represent an inherently diverse user population. CS1 and CS3 demonstrate how tabletop interfaces can support different user groups working with biological information. The study conducted in CS1 also shows how tabletops can support shared problem solving.

As discussed above, the visibility of actions and egalitarian input afforded by TEI systems can help to support collaboration [14, 20]. We believe the shared displays and multi-user input common in many TEI systems can help in communicating ideas between diverse users and in supporting the development of shared mental models, an important aspect of collaborative problem solving. This is crucial for learning, as in CS1, and for supporting collaboration between users with different expertise, as in CS2 and CS3. While TEI seems to have the potential to support collaborative scientific practices, much work is still needed to understand how it can best support the needs of the diverse stakeholders in the genomics community.

Area #5: Managing varied timescales

The case studies illustrate how research and collaboration timelines can range from one semester, as in CS1, to many years, as in the consortia of CS2 or the lab of CS3. Similarly, given the size and scale of genomics data, computational processing times can range from seconds to weeks or months. Maintaining awareness of and access to large data repositories over long periods of time (e.g. years) can hold transformative impact in genomics.

As discussed in [43], milliseconds and minutes are well understood to HCI, but insufficient for engaging longer interaction cycles. In human and cultural terms, heirlooms are passed down across generations, and some buildings maintained for thousands of years. TEI design, with its interweaving of the digital and physical, has the potential to address these varied timescales, and can investigate complementary interaction approaches, materials, and construction methods that address the varied informational, computational, and human timescales of genomics research.

CONCLUSION

We have provided a characterization of the genomics domain, raised challenges and questions for TEI research, and discussed lessons learned from three case studies. We have described potential areas in which TEI can contribute to the practice of genomics and also advance its own theories and methods. We expect effort in this space to also build on and contribute to other related fields, including CSCW, information visualization, and others. The interface between interdisciplinary fields is an exciting space; a vast region in which researchers can set off on wild explorations and forge completely new paths. These paths are not clearly determined; many open questions and challenges remain. We have begun to sketch some possible prospects and directions, and hope to have inspired both researchers and broader audiences to engage the challenges and pursue impact in the art and science of genomic interaction design.

ACKNOWLEDGMENTS

We thank Consuelo Valdes (Wellesley College) and Andy Wu (Georgia Tech). This work has been partially funded by NSF IIS-1017693, DRL-097394084, and CNS-1126739.

REFERENCES

1. Anderson, M. L. 2003. Embodied cognition: A field guide. *Artificial Intelligence*, 149: 91-130
2. Arnstein, L., Hung, C.-Y., Franza, R., Zhou, Q.H., Borriello, G,Consolvo, S., Su, J. 2002. Labscape: A Smart Environment for the Cell Biology Laboratory. In IEEE Pervasive Computing Magazine, 1(3). 13-2.
3. Bolchini, D., Finkelstein, A., Perrone, V., Nagl, S. 2009. Better bioinformatics through usability analysis, *Bioinformatics*, 25(3), 406-12, February 2009.
4. Brooks, F. P., Ouh-Young, M., Batter, J. J., Jerome Kilpatrick, P. 1990. Project GROPE: haptic displays for scientific visualization. In *Proc. of SIGGRAPH'90*, 177-185.
5. Bourzak, K. 2008. Biologists Enlist Online Gamers, *Technology Review*, MIT Press.
6. Carlson R.H. 2010. *Biology is Technology*, MIT Press.
7. Chin L, Hahn WC, Getz G, Meyerson M. 2011. Making sense of cancer genomic data. *Genes Dev*. 25:534 –555.
8. Fjeld, M., Fredriksson, J., Ejdestig, M., Duca, F., Bötschi, K., Voegtli, B., Juchli, P. 2007. Tangible user interface for chemistry education: comparative evaluation and re-design. *Proc. CHI'07*, ACM, 805-808.
9. Gillet, A., Sanner, M., Stoffler, D., Olson, A. 2005. Tangible augmented interfaces for structural molecular biology. *IEEE Comp. Graphics & Applications*, 25(2): 13-17.
10. Gray, J., Liu, D., Nieto-Santisteban, M., et al. 2005 Scientific Data Management in the Coming Decade. *Proc. SIGMOD'05*.
11. Han K., M. Konkel, J. Xing, et al. 2007. Mode and tempo of Old World monkey retrotransposon evolution: a glimpse through the Rhesus macaque genome. *Science*, v316.
12. Horn, M.S., Tobiasz, M., Shen, C. 2009. Visualizing Biodiversity with Voronoi Treemaps. *Proc. Sixth International Symposium on Voronoi Diagrams in Science and Engineering*, Copenhagen, Denmark. June 23-26, 2009.
13. Hornecker, E., Jacob, R.J.K., Hummels, C., Ullmer, B., Schmidt, A., van den Hoven, E., and Mazalek, A. 2008. TEI

Goes On: Tangible and Embedded Interaction, *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 91-96.

14. Hornecker, E., Marshall, P., Dalton, N.S., Rogers, Y. 2008. Collaboration and Interference: Awareness with Mice or Touch Input. *Proc. ACM CSCW Conference*.
15. Kuznetsov, S., Taylor, A.S., Paulos, E., DiSalvo, C., Hirsch, T. 2012. (DIY)biology and opportunities for HCI. *Proc. DIS '12*.
16. Locke, D., Hillier, L., Warren, W., et al. 2011. Comparative and demographic analysis of orangutan genomes. *Nature*, 469:529-533.
17. Mackay W.E, Pothier G, Letondal C, Boegh K, Sorensen H.E., 2002. The missing link: augmenting biology laboratory notebooks. *Proc. UIST '02*.
18. Marcus, A.D. 2011. Family Pioneers in Exploration of the Genome, *WSJ Health*, September 2011.
19. Mardis, E. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24(3), 133-141.
20. Marshall, P., Hornecker, E., Morris, R., Dalton, N., Rogers, Y. 2008. When the Fingers Do the Talking: a Study of Group Participation with Varying Constraints to a Tabletop Interface. *Proc. Tabletop 08*, 37-44.
21. Mirel, B. 2009. Supporting cognition in systems biology analysis: findings on users' processes and design implications, *J Biomed Discov Collab*, 27(2),153-155, February 2009.
22. Nersessian, N. J. 2002. The cognitive basis of model-based reasoning in science. *The cognitive basis of science*. Carruthers, Stich, Siegal, Eds. Cambridge U. Press: 133-153.
23. Nersessian, N. J. 2008. *Creating scientific concepts*. Cambridge, MA, MIT Press.
24. Nielsen C., et al. 2010. Visualizing genomes: techniques and challenges. *Nat. Methods*. 7:S5–S15.
25. Okada, T., Simon, H.A. 1997. Collaborative discovery in a scientific domain. *Cognitive Science*, 21 (2), 109-146.
26. Patten, J., Ishii, H. 2000. A Comparison of Spatial Organization Strategies in Graphical and Tangible User Interfaces, *Proc. DARE'00*, 41-50.
27. Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. The Rhesus macaque genome sequence informs biomedical and evolutionary analyses. *Science*, v316, 222-234, April 13. (Cover article.)
28. Rittle-Johnson, B. Star, J. R. 2007. Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3): 561-574.
29. Ryall, K., Forlines, C., Shen, C., Morris, M.R., Everitt, K. 2006. Experiences with and Observations of Direct-Touch Tabletops. *Proc. IEEE Tabletop'06*.
30. Saraiya, P., North, C., Duca, K. 2005. An insight-based methodology for evaluating bioinformatics visualizations, *Visualization and Computer Graphics*, 11(4), 443-456.
31. Schkolne, S., Ishii, H., Schroder, P. 2004. Immersive design of DNA molecules with a tangible interface, *Visualization*, IEEE, 227-234.
32. Schneider, B., Strait, M., Muller, L., Elfenbein, S., Shaer, O., Shen, C. 2012. Phylo-Genie: engaging students in collaborative 'tree-thinking' through tabletop techniques. *Proc. CHI'12*.
33. Scott, S. D., Grant, K. D., Mandryk, R. L. 2003. System guidelines for co-located, collaborative work on a tabletop display. *Proc. ECSCW'03*.
34. Shaer, O., Hornecker, E. 2010. Tangible User Interfaces: Past, Present, and Future Directions, *Foundations and Trends in Human-Computer Interaction*, 3(1-2), April 2010.
35. Shaer, O., Kol, G., Strait, M., Fan, C., Grevet, C., Elfenbein, S. 2010. G-nome surfer: a tabletop interface for collaborative exploration of genomic data. *Proc. CHI'10*.
36. Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., Wang, H. 2011. Enhancing Genomic Learning through Tabletop Interaction. *Proc. CHI'11*.
37. Shaer, O., Strait, M., Valdes, C., Wang, H., Feng, T., Lintz, M., Ferreirae, M. Grote, C., Tempel, K., Liu, S. 2012. The Design, Development, and Deployment of a Tabletop Interface for Collaborative Exploration of Genomic Data, *International Journal of Human-Computer Studies*.
38. Singer, S.R., Hilton M.L., Schweingruber, H.A. 2005. America's lab report: investigations in high school science, *National Research Council*.
39. Tabard A, Mackay W.E., Eastmond, E. 2008. From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. *Proc. CSCW '08*.
40. Tabard, A., Hincapié-Ramos, J-D., Esbensen, M., Bardram, J.E. 2011. The eLabBench: an interactive tabletop system for the biology laboratory. *Proc. ITS '11*.
41. Tang, A., Neustaedter, C., Greenberg, S. 2006. Videoarms: Embodiments for mixed presence groupware. *Proc. HCI '06*.
42. Tuddenham, P., Robinson, P. 2007. Distributed tabletops: Supporting remote and mixed-presence tabletop collaboration. *Proc. TABLETOP '07*.
43. Ullmer, B. 2012. Entangling space, form, light, time, computational STEAM, and cultural artifacts. *Interactions*, 19(4): 32-39.
44. Ullmer, B., Kim, E., Kilian, A., Gray, S., Ishii, H. 2001. Strata/ICC: physical models as computational interfaces. *CHI '01 extended abstracts*.
45. Veretnik, S., Fink, J., Bourne, P. 2008. Computational biology resources lack persistence and usability, *PLoS computational biology*, 7(4), July 2008.
46. Wigdor, D., Jiang, H., Forlines, C., Borkin, M., Shen, C. 2009. WeSpace: The Design, Development and Deployment of a Walk-Up and Share Multi-Surface Visual Collaboration System. *Proc. CHI '09*, Boston, MA.
47. Wilson, M. 2002. Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9(4): 625-636.
48. Wilson, S. 2002. *Information arts: Intersections of art, science, and technology*. Cambridge, MA, MIT Press.
49. Wu, A., Caspary, E., Yim, J.-B., Mazalek, A., Chandrasekharan, S., Nersessian, N. J. 2011. Kinesthetic pathways: A tabletop visualization to support discovery in systems biology. *Proc. Creativity and Cognition*, 21-30.
50. Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J., Paepcke, A. 2006. ButterflyNet: a mobile capture and access system for field biology research, *Proc. CHI'06*, 571–580.