Management, dissemination, and training core

Specific Aims

Aim 1. Provide effective scientific, project, and personnel management.

Aim 2. Broadly disseminate the software tools, libraries, source code and data.

Aim 3. Train users with online documentation, videos, workshops, courses, email and social media.

Research Strategy

Aim 1: Provide effective scientific, project, and personnel management

Significance

We have consistently placed a high value on the benefits of a good management team and adherence to solid project processes and best practices. The scientific guidance of our PIs, Scientific Advisory Board (SAB), and colleagues focuses our work on efforts that best support our research community, while steering us in the right direction to face upcoming challenges. Our strong project management keeps our diverse engineering team working efficiently and effectively, while staying attuned to our funding agency's expectations and the needs of our resource users. Our personnel managers are ranked among the best at UCSC, maintaining a high level of performance from our staff in tandem with keeping them happy and engaged in their work.

Innovation

We intend to retain our current management processes and tools, which have proven quite successful, while remaining cognizant of new tools and ideas that might benefit our group. We plan to review the areas of expertise covered by our current SAB and identify potential new advisors that bring additional knowledge to the group.

Approach

Genome Browser organizational structure and staff responsibilities (see Fig. 1)

Management

(* = Key Personnel)

* <u>W. James Kent, Ph.D.</u> – Principal Investigator, Director UCSC Genome Browser. Sets overall direction and priorities of project. Software architect.

* <u>David Haussler, Ph.D.</u> – Co-Investigator, Distinguished Professor Biomolecular Engineering, HHMI Investigator. Helps set scientific direction and priorities. Represents project to outside scientific community. Fundraising. Interfaces with consortia.

* <u>Robert Kuhn, Ph.D.</u> – Associate Director and Outreach, UCSC Genome Browser. Helps set scientific direction and priorities. Directs outreach and training programs. Interfaces with other database providers. Licensing support.

<u>Ann Zweig</u> – Senior Project & Engineering Manager. Organizes workflow overall into discrete tasks and assigns tasks to individual staff. Manages software developers. Writes and edits progress reports, grants, and scientific papers.

<u>Brian Lee</u> – Quality Assurance Manager. Manages Genome Browser QA program. Trains and reviews work of QA engineers. Leads internal documentation effort.

Software Development

<u>Angela Hinrichs</u> – Senior Software Developer, Technical Lead. Human variation data and tools. JavaScript and user interface.

<u>Kate Rosenbloom</u> – Senior Software Developer. GTEx data and integration. Protein data.

Hiram Clawson – Senior Software Developer and Data Integrator. Comparative genomics. Assemblies and annotation tracks.

Galt Barber – Senior Software Developer. Network and parallel computing. Data integration tools development. MySQL database administration. Multi-region view. User security. Speed optimization.

Brian Raney, Ph.D. – Senior Software Developer. Display and graphics programming. Track hubs.

Jonathan Casper – Software Developer and Data Integrator. JavaScript and user interface.

Max Haeussler, Ph.D. – Assistant Research Scientist. Provide scientific guidance to project. Software developer. Integration with external tools developers. Text mining.

Quality Assurance

Matthew Speir – Lead Quality Assurance Engineer. Development of database-oriented test scripts. Trains new QA staff.

Luvina Guruvadoo – Quality Assurance and User Support. Software testing and third-party data validation. Manages mailing lists.

Christopher Lee – Quality Assurance and User Support. Leads QA on user products.

Cath Tyner – Quality Assurance and User Support. Software testing and third-party data validation. Reviews user interface changes.

Chris Villarreal – Quality Assurance and User Support. Software testing and third-party data validation. Performs biological spot-checking on data.

Jairo Navarro – Quality Assurance and User Support. Software testing and third-party data validation. Reviews user interface changes.

Systems Administration

Erich Weiler – Genomics Systems Architect. Design, installation, and maintenance of compute, storage, and network systems. Design of remote clusters.

Haifang Telc – Computer Systems Administrator. Installation, configuration and maintenance of compute, storage, and network systems. Integrates UCSC and Genome Browser network and services.

Administration

Nadine Gassner, Ph.D. – Grant Administrator, Genomics Institute. Oversees or coordinates administrative functions associated with the UCSC Genome Browser, including proposals, reporting, financial management, resources and personnel. Ensures coordination with other large institute grant projects.

Ned LeBlond – Grant and Program Support Analyst. Assists Grant Administrator.

Gina Bentley – Financial Manager. Purchasing, reimbursements, travel advances, and other financial transactions. Ensures appropriate use of funds in compliance with federal and UC guidelines.

Lilia Sandoval – Human Resources. New hires. Personnel support.

Rochelle Fuller – Grant Support Administrator. Technical support on administrative software. Prepares final versions of reports, letters, and forms. Special projects.

Work-study students – Basic clerical functions associated with Genome Browser staff and space. Administrative support.
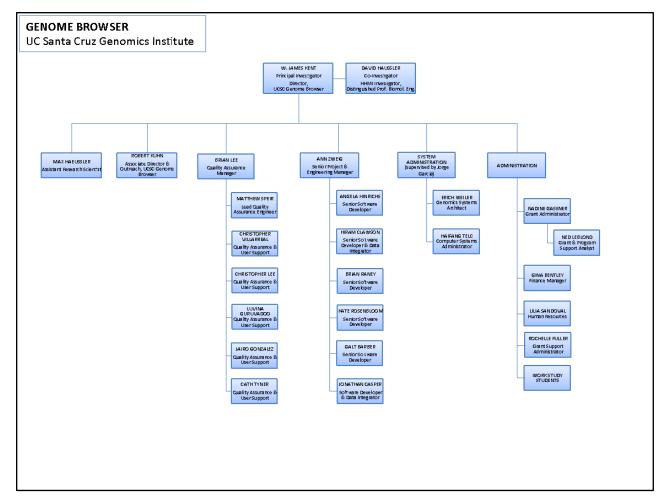
GENOME BROWSER
UC Santa Cruz Genomics Institute

W. JAMES KENT
Principal Investigator
Director,
UCSC Genome Browser

DAVID HAUSSLER
Co-Investigator
HHMI Investigator,
Distinguished Prof. Biomol. Eng.

MAX HAEUSSLER
Assistant Research Scientist

ROBERT KUHN
Associate Director &
Outreach, UCSC Genome
Browser

BRIAN LEE
Quality Assurance
Manager

ANN ZWEIG
Senior Project &
Engineering Manager

SYSTEM
ADMINISTRATION
(supervised by Jorge
Garcia)

ADMINISTRATION

MATTHEW SPEIR
Lead Quality
Assurance Engineer

ANGELA HINRICHS
Senior Software
Developer

ERICH WEILER
Genomics Systems
Architect

NADINE GASSNER
Grant Administrator

CHRISTOPHER
VILLARREAL
Quality Assurance &
User Support

HIRAM CLAWSON
Senior Software
Developer & Data
Integrator

HAIFANG TELC
Computer Systems
Administrator

NED LEBLOND
Grant & Program
Support Analyst

CHRISTOPHER LEE
Quality Assurance &
User Support

BRIAN RANEY
Senior Software
Developer

GINA BENTLEY
Finance Manager

LUVINA
GURUVADOO
Quality Assurance &
User Support

KATE ROSENBLOOM
Senior Software
Developer

LILIA SANDOVAL
Human Resources

JAIRO GONZALEZ
Quality Assurance &
User Support

GALT BARBER
Senior Software
Developer

ROCHELLE FULLER
Grant Support
Administrator

CATH TYNER
Quality Assurance &
User Support

JONATHAN CASPER
Software Developer
& Data Integrator

WORKSTUDY
STUDENTS

Figure 1. Genome Browser organizational structure and staff responsibilities.

Progress reporting

Progress will be reported as required by the funding agency. At a minimum, we will provide an annual report of accomplishments and plans.

Scientific Advisory Board

The UCSC Genome Browser project has an active Scientific Advisory Board (SAB) that currently consists of six members: Mary-Claire King (University of Washington), Tim Hubbard (King's College London), Aravinda Chakravarti (Johns Hopkins University), Robert Waterston (University of Washington), Ross Hardison (Penn State University) and Barbara Wold (CalTech).

Key personnel from the project team meet with the SAB and our NHGRI Program Director, currently Valentina Di Francesco, every 12-15 months to present major accomplishments and seek the SAB's advice on priorities, problems, and future directions. Efforts are made to make the meetings as inclusive and efficient as possible: agendas and supplemental materials are provided to attendees in advance, and videoconferencing facilities accommodate members who are unable to attend in person. Following each meeting, a summary of the key decisions and action items is distributed to attendees, and items that directly affect the priorities of the Genome Browser project team are incorporated into the project planning.

Periodically we review the areas of expertise of our board members, identify areas in which we'd like more representation, and then target potential new board members in these fields.

Aim 2: Broadly disseminate the software tools, libraries, source code and data

Significance

The Genome Browser website at https://genome.ucsc.edu/ is the primary vehicle for disseminating the resources produced by our team and the data contributed by a large group of collaborators. Widely known and used within the genomics community, our website averages 1.2 million hits each day (see Fig. 2, Overall component) (1–5).

The presence of a web application or a large set of files on the Internet is not by itself sufficient for people to find the resource. Over the last sixteen years we have developed extensive dissemination channels for distributing our software, data and usage information. The awareness of our resource is widespread throughout the scientific community, evidenced by its citation record in the literature, links from collaborating websites, and frequent mention in presentations at scientific meetings (6–11). We also ensure that the Genome Browser website is accessible to web search robots in a regulated manner that avoids performance degradation.

Innovation

Despite the continued growing awareness of the Genome Browser within the genomics community, it is important that we simultaneously reach new users while encouraging our long-time users to explore new data sets and take advantage of new features. We support all users in learning more about our website in three major ways: documentation, user support, and user training. We describe our intended plans in these areas in Aim 3.

Approach

The Genome Browser website provides convenient centralized access to our complete set of graphical tools for visualizing and manipulating the genomic data in our databases, along with supporting documentation and access to downloadable data. Our public MySQL server (http://genome-mysql.cse.ucsc.edu) and a variety of Internet file transfer protocols provide access to the data in bulk and computer-readable form. Third-party data is made available on our site following a check to ensure that it meets our quality standards. Alternatively, users can download both the software and source code through the Genome Browser online store at https://genome-store.ucsc.edu/. Data downloads are freely available to all users (with a few specific restrictions imposed by contributing institutions). Software and source code downloaded by commercial users require a license, but are freely available for non-commercial and academic use.

In addition to providing access to software tools and data, the website offers a contact page, a newsfeed that is actively updated with announcements of releases of new data and software, an extensive collection of online documentation and tutorials, and links to a large number of other websites offering genomics tools that complement those offered by the Genome Browser (12–24). Many external mirror sites have set up scripts that sync with the browser ftp site on a regular basis. The website provides detailed instructions for setting up a Genome Browser mirror, as well as a description of the API for those individuals who wish to programmatically interface with the browser through such mechanisms as hubs and custom tracks. We provide several mechanisms through which users can provide suggestions and feedback for improving our website and processes (such as a user support mailing list and a suggestion box), and incorporate these ideas into our ongoing website and tools development when feasible.

Aim 3: Train users with online documentation, videos, workshops, courses, email and social media

Significance

As the volume and complexity of genomic data continues to increase at an unprecedented rate, we regularly introduce new data types, visualization features, and tools to the Genome Browser to help users browse and analyze the data (25–30). In tandem with the dissemination efforts described in Aim 2, it is essential that resources such as ours also invest in training to help our users effectively apply our tools and data sets to their work. As our user base continues to grow and diversify, new users need basic instruction on how to best use the resource for their particular research requirements, and advanced power users (who in some cases have

been using the browser for more than a decade) enjoy learning about the newest features and how to apply them to their best advantage.

Innovation

Since the inception of the UCSC Genome Browser sixteen years ago, we have provided training in diverse formats (Appendix 3). Robert Kuhn, Associate Director of the UCSC Genome Browser project, currently spends the majority of his time supporting the outreach and training needs of our users. Through his close contact with our user base, he is well suited to represent their needs when providing input to the scientific direction of the project.

Our existing training and outreach strategy includes a combination of in-person workshops at institutions and scientific meetings, as well as online documentation and videos, a user support help desk, and social media content. The in-person workshops, while labor-intensive for the trainer, are an excellent way to reach users who want a specific deep understanding of the Genome Browser. We fine-tune the content to meet the exact needs of the attendees, and workshops tend to help build a local base of committed power users who, in turn, are evangelists for the Genome Browser (Appendix 4). In contrast, our online content, while more generic to our user base, allows us to efficiently inform a much broader audience with lower staff effort. Our public help desk support taps the advantages of both: while correspondences are geared towards the needs of a specific individual, all subscribers to the mailing list benefit from the information exchange.

The Genome Browser SAB has recommended that we consider designing a MOOC (massive open online course) as a means for combining the advantages of in-person training with the efficient mass outreach capabilities of online content. We have also considered offering an annual UCSC-based short course (of approximately 3-day duration) that would provide attendees with a deeper level of training than our current workshops, which are typically limited to a day or two in length. However, neither of these undertakings is feasible given our current staffing effort.

Approach

As we have done with great success for many years, we plan to continue offering workshops at institutions in exchange for payment to meet our costs: flights, lodging, and a small flat rate to cover incidentals. We do not consider it feasible to charge enough to cover the salary of the trainer in an academic environment, though this is often possible in commercial environments. To date, the workshops that we have been invited to present at institutions, as well as those at scientific meetings, have been well attended and quite successful. We have received repeat invitations to present at several conferences of specific interest to our target audience, such as American Society of Human Genetics (ASHG), European Society of Human Genetics (ESHG), and Plant and Animal Genomics (PAG). Many attendees who describe themselves as "advanced browser users" have reported that they learned new information about the browser while at the workshops, making it well worth the amount of time spent. During the next 5-year period we plan to continue our practice of giving workshops at scientific meetings and institutions.

The Genome Browser website offers a large collection of online help documentation, tutorials, and user guides. Each new feature release is accompanied by user help documentation in an appropriate format, such as a news announcement, a short video, and/or a full-featured user guide. Additionally, releases of new features and data sets are announced on our news feed, twitter feed and Facebook page, and topics that deserve a longer format are showcased in our blog. To date, we have produced nine short video tutorials accessible from the Genome Browser YouTube channel that help users learn how to perform common functions and use the browser to answer basic biological questions. We gather ideas for new video tutorials from users at workshops and from questions asked repeatedly on our support mailing list, and typically highlight topics that are broadly useful but might be difficult to discover on the site. We plan to maintain our current strategy for documentation, social media content, and video production during the next grant cycle.

We also intend to continue our current level of help-desk staffing during the upcoming grant cycle. Our help desk spends a minimum of eight hours per day responding to email questions from users (more than 80 email threads per month), which have grown in quantity, complexity and scope over the years. In addition to new users in need of basic guidance, many of our more advanced users are pushing the limits of the resource and ask questions commensurate with these complicated uses. Basic questions from beginners lead us to review

our documentation to see if something is unclear, possibly resulting in the creation of a short video tutorial to which we can refer future users. Suggestions from power users are fed back into our internal feature-tracking system for possible implementation. In addition to our general technical support mailing list, we offer assistance to thousands of users around the world who have downloaded and installed their own copy of the Genome Browser website and tools to mirror our site.

The user support and training functions provide us with an excellent opportunity to collect feedback from our users. The suggestion box on our website, which currently garners 30-40 inputs per year, provides a venue for users to recommend new features, novel data sets, and improvements to increase the usability and value of the Genome Browser. The Project Manager replies to each suggestion, and creates tickets in our internal request tracking system for those that will likely be of general use to a broad audience. We also seek user feedback through regular surveys posted on the website and through lab visits where we observe people using our research tools during the course of their normal work (Appendix 2). Our on-site training sessions are also a rich source of ideas. We plan to continue support for the suggestion box and periodic surveys in upcoming years.