

Training and Outreach

The training and outreach program has two main objectives: reaching new users and supporting existing users. The former is best achieved by in-person workshops, the latter both in person and by online resources such as help pages and video tutorials. In the course of attending a 90-minute, half- or two-day workshop at their institution, novices in genomics can learn the scope of the Browser toolset and become quite adept at using it. Furthermore, the Browser team learns from being in the room with our users, whether expert or beginners, such as what they need and where we can improve.

Staff professional development

Our staff participated in the following development opportunities during the reporting period, attending classes at UCSC:

- Most of our staff took UCSC-required Cyber Security training, Sexual Harassment Prevention training and Conflict of Interest training. All members of the staff have completed the NIH training course “Protecting Human Research Participants”.
- Staff members regularly attend seminars at UCSC on topics relevant to our work -- both bioinformatics and molecular biology.
- Several staff members audited or took for credit regular UCSC courses:
 - Matt Speir, Christopher Lee: Introduction to data science in Python.
 - Cath Tyner, Christopher Lee, Christopher Villarreal, Jairo Navarro Gonzalez: Genomes.
- During the past year, the Genome Browser project provided research training and professional development opportunities for an undergraduate student:
 - David Gibson worked on audio/video editing for YouTube tutorials.

Workshop trainings and meeting appearances

- This year we continued to offer trainings in the use of the Genome Browser and its attendant tools, including, where appropriate, instruction in command-line approaches to sequencing-file manipulation, the display of user-generated sequencing results in the Browser, and comparative genomics datasets. In general, the trainings have been requested more by clinicians and molecular biology researchers than by bioinformatics professionals, who tend to use downloads and write their own scripts, rather than use the web-based user interface. The trainings were conducted in a variety of venues, including international meetings of scientific societies, and at individual institutions over the course of 40 full or partial days of instruction (**Table B.4.1.**).
- During the funding period, presentations were made to more than 500 attendees at national or international meetings of the American Society of Human Genetics (ASHG), European Society of Human Genetics (ESHG), Korean Genetics Organization (KOGO), American Society of Bone and Mineral Research (ASBMR) and Plant and Animal Genomics (PAG). The large-meeting format, while

shorter than that of institutional workshops, allows us to reach people from a large number of institutions at once. These workshops afforded an opportunity to learn what our users need and how we can make our tools more intuitive and easy to use.

- Our presentations at large meeting have often resulted in invitations to visit individual institutions for longer workshops, where we can seed a core local user group with deeper instruction. Such intensive workshops, of half- to two-days duration, were given to 1000 scientists at 25 locations this year.
- In addition, we were invited to present the Browser in two training courses this year, one the 58th annual McKusick Mammalian Genetics course sponsored by JAX Lab and Johns Hopkins in Bar Harbor, ME, the other, on Variant Effect Prediction, sponsored by Human Genome Organisation (HUGO) and the Human Variome Society (HGVS) at Heraklion, Greece.
- We have presented three posters at meetings: ASHG, Hinxton Rare Disease meeting, Cold Spring Harbor Lab CRISPR meeting.
- We have presented Browser content on conference calls for GA4GH and GTE_x (2) Consortia as well as an hour-long webinar for curators at Ambry Genetics.
- Our training seminars are well received, and our post-workshop surveys indicate that even regular users discovered new information about the Browser's functionality.
- Since we began posting our upcoming workshop schedule on the site, many of our workshop hosts have been contacted by people from out of town about attending. Attendees have traveled sometimes several hundred miles including an overnight stay.
- Onsite training classes also provide valuable insight into the data and software/display needs of the user community. E.g., Our re-engineering of the RefSeq and UniProt tracks came about after users asked about them in workshops.
- We have learned in our workshops that people in the room may not raise their hands with a question, but they will ask when the instructor circulates nearby during the problem-solving portion of a workshop. Similarly, people who will not make a request by contacting the mailing list (e.g., for Linkage Disequilibrium data on hg19) are less shy about asking when they meet us at meetings or workshops.
- OpenHelix (OH) continues to maintain three online tutorials about the Genome Browser, one of which was updated this year. Views of their Genome Browser videos on the OH site number in the 1000s. They also maintain and distribute Quick Reference Cards (QRCs, updated this year) covering the Genome Browser and Table Browser.

Table B.4.1. Presentations. Training workshops and public presentations presented during this reporting period. Institutional locations are one- or two-day workshops, those at national or international meetings are typically 90 minutes. These presentations were offered by Robert Kuhn, Associate Director of the Browser project unless otherwise indicated -- Matt Speir, Brian Lee, Maximilian Haeussler.

Date	Workshop or Presentation
July 2016	UCSC COSMOS Program
	JAX Lab Mammalian Genetics course, Bar Harbor, ME
August 2016	Cold Spring Harbor Labs CRISPR meeting (Haeussler)
September 2016	UC San Francisco
	Tuskegee University, AL
	University of Colorado Denver Medical Campus

October 2016	Penn State University
	European Bioinformatics Institute, Hinxton, UK
	American Society of Human Genetics, Vancouver, BC (+ Speir)
November 2016	Variant Effect Prediction course, Heraklion, Greece (3 talks)
December 2016	UCSC. Research mentorship workshop
January 2017	Plant and Animal Genomics, San Diego, CA (+Lee)
	Aptos HS, Aptos, CA
February 2017	Catholic University, Seoul, Korea
	KT Education Center, Daejeon, Korea
	Korean Genome Organization symposium, Vivaldi Park, Korea (2 talks)
	Mt. Sinai Medical School, NYC
	Regeneron, Inc, Tarrytown, NY
April 2017	Howard University, Washington, DC
	University of Virginia, Charlottesville
	RNA-seq meeting, San Francisco, CA
May 2017	European Society of Human Genetics, Copenhagen, Denmark
June 2017	Southwestern University, Georgetown, TX
	Washington University, St. Louis, MO
	University of Texas San Antonio, TX
	UCSC Summer workshop
July 2017	Ambry Genetics (online Webinar)
	University of Pennsylvania, Philadelphia
	JAX Lab Mammalian Genetics course, Bar Harbor, ME
August 2017	UC San Francisco

Browser training video tutorials

- This year we have released three new video tutorials detailing specific Browser tasks for display on our YouTube channel, bringing the total to fourteen (see: <https://www.youtube.com/channel/UCQnUJepyNOW0p8s2otX4RYQ/videos>). Each is released with an index and a transcript. There are two more in the final stages of production. These videos focus on specific features (such as the three types of Multi-region display, including the popular exon-only display), but they also incorporate navigation features that casual users are not likely to have discovered. In that way, we expose our users to methods that will expand their capabilities with the Browser. The channel has 605 subscribers. Our fourteen videos were viewed 11,700 times during the reporting period.

Blog posts and MLQs

- We maintain a low-frequency blog (<http://genome.ucsc.edu/blog/>) to announce new features and to describe certain aspects of the Browser. During the reporting period, we released five blog posts

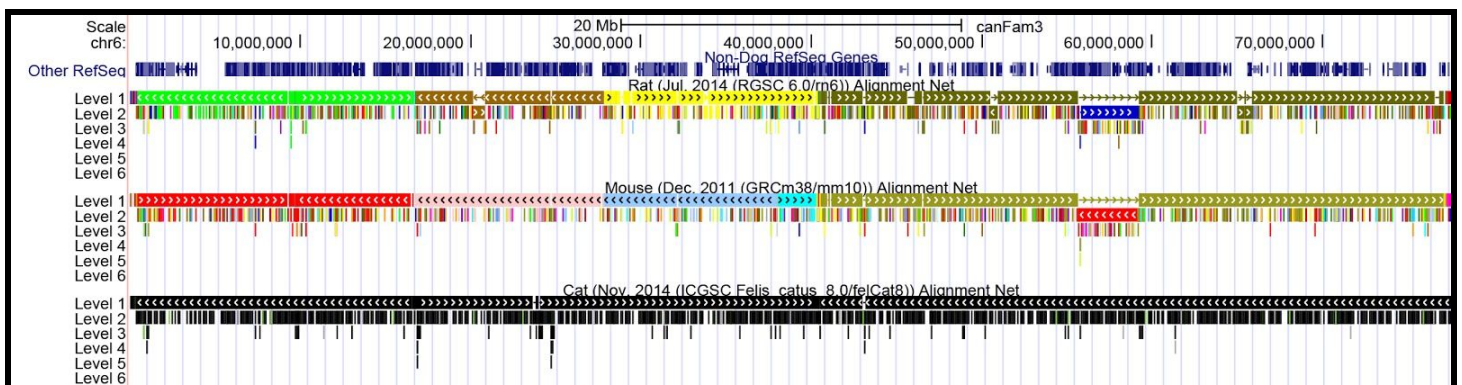
describing the Variant Annotation Integrator, Track Hubs, the re-engineering of our RefSeq display, zero-based coordinate data storage and our GTEx data display.

- Our mailing lists continue to attract many questions and suggestions. During the year, we responded to more than 1000 threads, many of which included multiple back-and-forth communications with the user. Our mailings are public and searchable, so that subsequent users with a similar question can find our replies. We also have a private mailing list for those with confidential inquiries. Our mailing lists supplement our on-location workshops in providing a source of information about features that our users would like to see.
- Additionally, our staff occasionally replies (fewer than 100) to UCSC-related questions on biostars.org bioinformatics public forum.
- We also receive suggestions from our users through our online Suggestion Box. This year there were 39 suggestions from users.

Session gallery: Highlighting Genome Browser features

- Last year we added a session gallery (<http://genome.ucsc.edu/goldenPath/help/sessions.html>) featuring sample sessions that highlight topics of interest in molecular biology education, such as the display of coding and wobble bases, alt-splicing, evolution, variation and disease as well as topics derived from commonly asked questions on the Browser user mailing list and feedback from onsite Browser training workshops.
- We have made it possible for users to make their sessions public, visible to all other users. During the reporting period, more than 200 publicly-shared sessions have been created out of the at least 40,000 new sessions created.

Figure B.4.1. Example shared educational session. In this session from the Genome Browser gallery, the chainNet tracks on chr6 of the dog genome illustrate the close synteny between the dog and cat (lower black track) and their differences with the rat-mouse pair (upper two tracks, multiple colors indicating different chromosomes), themselves quite similar to each other. Illustrates chromosome rearrangements and the evolutionary concept of shared ancestry.



Support for NAR authors

During the year, it came to our attention that the Journal, *Nucleic Acids Research*, has begun to require that authors reporting on large data sets not only make the data available, but that they supply to the reviewers a link to a Saved Session on the UCSC Genome Browser: The journal requires that for papers reporting whole genome expression or sequencing analyses using ChIP-seq, RNA-seq or other types of whole genome sequence-based analyses, all data sets be deposited in GEO (or equivalent public repository) and accession numbers be provided in the manuscript. Private access links to the deposited data, AND to a UCSC genome browser session displaying the uploaded sequence tracks, must also be uploaded as unpublished datasets that can be accessed by reviewers. In support of these authors, we have documented the session construction process, in cooperation with the journal, and provided a link for the journal to offer authors:

<https://genome.ucsc.edu/goldenpath/help/hgSessionHelp.html#NAR>.

Track Hub versioning

For many years we have coordinated with the Ensembl group at EBI to standardize file formats, release data for the same genome assemblies, and collaborate in other areas where we can reduce confusion in the community. To this end, the Track Hub mechanism designed by UCSC has been adopted by Ensembl to allow users to display their data at the Ensembl Browser. During the reporting period, we collaborated with the Ensemble to update a versioning system that allows sites supporting Track Hubs to support a particular, well defined set of parameters in their hubs. User can then validate their hubs against the evolving version standards over time.

User Survey

In July, 2016, we queried our users via a survey to discover their demographics and interests and what features work for them or do not, and how we might improve the Browser to meet their needs. Among interesting results:

- We learned that 85% of respondents hold a post-baccalaureate degree, with 70% of those being PhD and/or MD.
- Sixty-five percent work at academic or non-profit locations.
- They are doing bioinformatics; whole-genome and whole-exome sequencing in equal measure; PCR, microarrays, and already in July 2016, CRISPR (15% of respondents).
- Users' interests are predominantly disease (non-cancer, cancer), cell biology and epigenetics, but range widely beyond that as well.
- Human and mouse are reported as the most commonly used assemblies, which matched information gleaned from our logs.
- Forty percent report regularly using Ensembl as well as UCSC; 26% are using IGV regularly.

- There were also a significant number of feature requests, many of which were incorporated into future plans as outlined in our grant renewal proposal.

Genome Browser mirror support

- The European mirror (genome-euro.ucsc.edu) in Bielefeld, Germany: This year we doubled the server's disk capacity to accommodate increases in our data footprint and established a public MySQL server. We encountered unexpected international logistical roadblocks that prevented an American institution from purchasing equipment through Dell for shipment to Germany, which delayed the project by many months. The faster server is now operational and the download server is finally in place and about to be enabled. Changes in the scripts on the Genome Browser in a Box (GBiB) product were necessary to fully implement this feature in an automatic fashion and are underway.
- The Asian mirror (genome-asia.ucsc.edu) at RIKEN in Japan: The server went live in July 2016. This is a fully functional copy of the Genome Browser maintained by UCSC. It is currently operational and we have enabled automatic redirect of Browser traffic originating in Asia to this mirror. All hardware for this mirror site was purchased directly by RIKEN. RIKEN and UCSC have signed a Memorandum of Understanding in which RIKEN agrees to provide electricity, cooling and hardware support for the machines. We provide software and Browser expertise in exchange for their making the mirror accessible to all of our users.

Program income to support future trainings

- Workshops at national meetings are typically given at our expense, though we obtain a waiver of registration fees that limits our expenses to the cost of travel.
- Institutional workshops continue to operate primarily on a "host pays" basis for plane and accommodation, and also generate program income.
- Our program income for this grant year totals \$12,800 in collected income for workshops completed during the reporting period. This amount covered 101% of the costs of travel incurred at UCSC for all Browser training purposes during the period. There is a balance of \$27,650 from previous years for future travel.

Genome Browser citations in the literature

- A survey of formal citations in the literature using Google Scholar indicates that the papers describing the Genome Browser and its tools have received more than 27,000 formal citations in the literature, 18% of which occurred in the past year (**Table B.4.2.**). This list does not include papers on which our staff have co-authorship that cover topics other than the Browser, such as papers about the Genome 10K project, the release of human or other genome assemblies, or bird evolution.
- The journal *The Scientist* recently reported on an analysis by Clarivate Analytics of the most-cited scientists and engineers in the world. Among the 1,644 U.S. scientists listed in the top 1% is included a significant number of our staff: Galt Barber, Hiram Clawson, Angie Hinrichs, Donna Karolchik, Jim Kent,

Robert Kuhn, Brian Raney, Kate Rosenbloom, Ann Zweig and former staff: Mark Diekhans, and Rachel Harte. Ref:

https://www.the-scientist.com/?articles.view/articleNo/50965/title/Clarivate-Ranks-Most-Cited-Researchers-of-2017/&utm_campaign=NEWSLETTER_TS_The-Scientist-Daily_2016

Table B.4.2. Formal citations of the Genome Browser and its tools in the literature.

Topic	Author, Year	Google 2017
BLAT	Kent, 2002	6,148
Genome Browser	Kent et al., 2002	5,609
Conservation	Siepel et al., 2005	2,406
Browser database	Karolchik et al, 2003	1,472
Table Browser	Karolchik et al., 2004	1,168
Threaded Blockset Aligner	Blanchette et al., 2004	1,088
Genome Browser update 2011	Fujita et al., 2011	1,042
Genome Browser update 2013	Meyer et al., 2013	641
Chain/Nets (evolution's cauldron)	Kent et al., 2003	624
Genome Browser update 2010	Rhead et al., 2010	609
Genome Browser update 2006	Hinrichs et al., 2006	564
Genome Browser update 2008	Karolchik et al., 2008	560
Genome Browser update 2014	Karolchik et al., 2014	557
ENCODE resources 2013 - 5-yr update*	Rosenbloom et al., 2013	494
Known Genes	Hsu et al., 2004	456
Genome Browser update 2015	Rosenbloom et al., 2015	453
Genome Browser update 2009	Kuhn et al., 2009	371
Genome Browser extensions & updates	Dreszer et al., 2011	345
Genome Browser update 2007	Kuhn et al., 2007	320
ENCODE whole-genome data*	Rosenbloom et al., 2009	245
BigWig and BigBed	Kent et al., 2010	242
28-way alignment	Miller et al., 2007	233
ENCODE resources 2012*	Rosenbloom et al., 2012	210
Browser and associated tools	Kuhn et al., 2013	203
Current Protocols	Karolchik et al., 2009	199

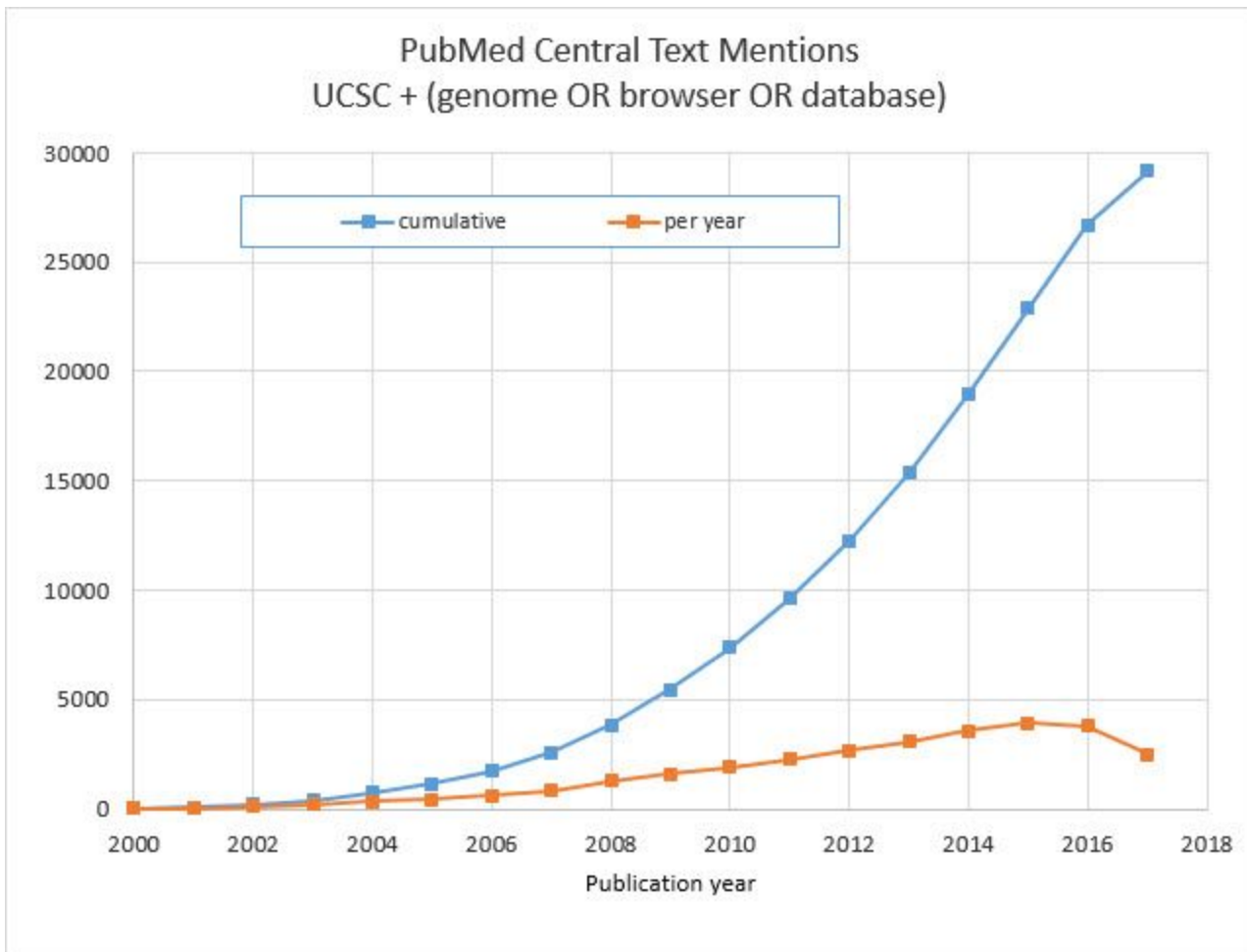
ENCODE resources 2011*	Raney et al., 2011	186
Genome Browser update 2016	Spier et al. 2016	175
Archaeal Browser*	Schneider et al., 2006	123
Track Hubs	Raney et al., 2013	119
Cancer Browser*	Zhu, et al., 2009	117
ENCODE resources 2007*	Thomas et al., 2007	96
Gene Sorter	Kent et al., 2005	72
Browser Tutorial	Zweig et al., 2008	56
Proteome Browser	Hsu et al., 2005	52
ENCODE 2016	Sloan et al 2016	50
Genome Browser update 2017	Tyner, et al. 2017	48
Current Protocols	Karolchik et al., 2011	45
Comparative Genomics with Browser	Karolchik et al., 2008	27
Biotechnology Annual Review - deep support	Mangan et al., 2008	25
Understanding Genome Browsing	Cline et al., 2009	23
Genomic Data Resources	Lathe, et al., 2008	22
GBiB	Haeussler et al., 2014	19
Current Protocols 2009	Mangan et al., 2009	16
Genome Space	Qu et al, 2016	13
GTEX*	GTEX Consortium, et al.	10
Current Protocols 2014	Mangan et al., 2014	9
Comparative Assembly Hubs	Nguyen et al., 2014	8
VAI and Data Integrator	Hinrichs et al. 2016	8
Ebola Portal	Haeussler et al., 2014	6
	total	27,584

* Resources not directly or fully funded by main Genome Browser grant, but featuring Genome Browser

- Many researchers who use the Genome Browser or database mention it in the text of the paper, but do not include a formal reference. By querying the text of papers in PubMed Central (PMC), we have identified more than 29,000 mentions of “UCSC AND (genome OR browser OR database)” in the portion of literature available as full text to PMC (**Fig. B.4.2.**). It is not clear what proportion of total publications are available to the search in PMC, but there are certainly papers citing the Browser that

are not available as full text in PMC. It appears that as the Browser becomes more widely used, it is being viewed more as a utility and is not referenced in the formal references. Nevertheless, both formal references and casual mentions in the text of publications continue to increase year over year..

Figure B.4.2. Text mentions in PubMed Central “UCSC + (Genome OR Browser OR database)”. Because PMC contains only those papers released for full-text public access, there is often a significant lag time between a paper’s release and its archival in PMC. For example, a significant number of papers published in 2015 did not reach PMC until 2016. Because of this, the last two time points in the “per year” portion of the graph are incomplete.



Genome Browser usage

General website usage

- Statistics from this grant year (gathered from our website logs):
 - We averaged ~47 million hits per month for our three sites combined: UCSC-based site (42 million), genome-euro (4 million) and genome-asia (1 million). Our website logs count a “hit” as any request from a single user’s computer whether it’s to a CGI, an image, an HTML page, etc,

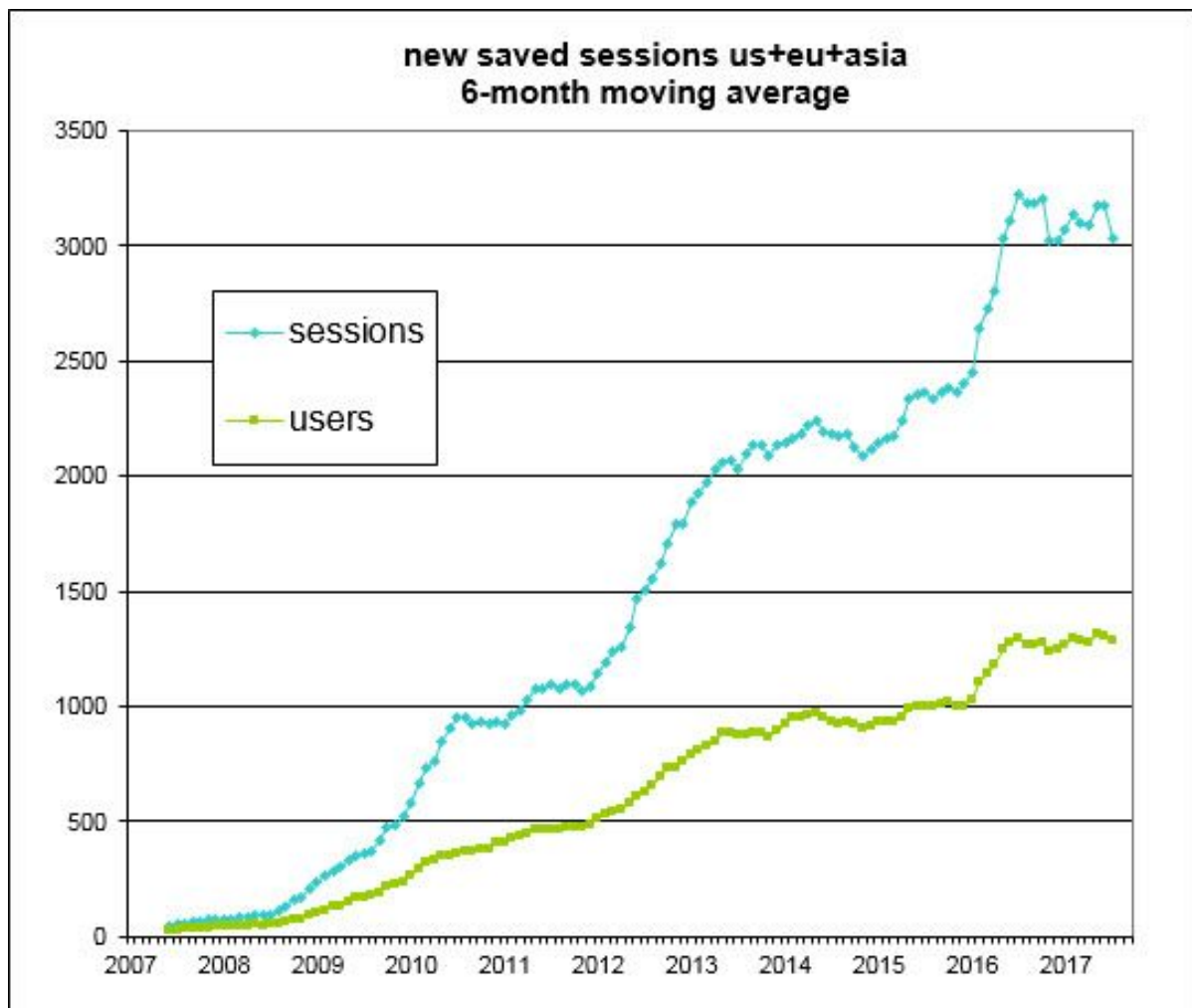
and are an overestimate of activity. These do not count any activity on mirrors, GBiB or GBiC, of course. Google Analytics data (below) tracks individual user sessions.

- These hits are from 175,000 IP addresses per month averaged over the grant year, up 4% over last year.
- 390 IP addresses downloaded and installed the GBiB product. We also counted an additional 180 locations showing active engagement with their installation by emptying trash, which resets the reporting mechanism and triggers a fresh ping to our logs. Active installations that do not empty their trash are not recorded.
- We have counted 24 installations of GBiC utility since its release in January, 2017
- Since activating Google Analytics (GA) for our CGIs (software web tools) on the Genome Browser website, we have collected the following interesting statistics:
 - We experience 'active' users at the rate of ~125,000 per month, with the average user logging 11 page views per visit. Note that Google Analytics are activated only on our CGIs; not on the rest of our pages (static pages). This explains the difference in count between the Google Analytics "users" and the "IP" count above.
 - We receive ~5 million page views per month with no upward or downward trend, except for seasonal fluctuations. 90% of the traffic is on the US site, 7% on the European site and 2% on the Asian site. GA counts a "hit" as a single visit to a CGI by a single user. Automated robots and spiders are not included in this value.
 - 39% of all browsing sessions are from the USA, 12% from China, 6% from the UK, 4% from Japan, and 4% from Germany. Apparently, our Chinese users are mostly using the US site. This may be because the redirect to the Asian site will only be active for a new user visiting the gateway page. Existing users already using the US site may not become aware of the site in Asia for some time. Statistics show that visits to the hgTracks cgi (main graphic page) exceed visits to hgGateway (where the redirect occurs) by a factor of 8.7 in a typical month.
- Our downloads servers are responding to 5 million requests per month from 30,000 unique IP addresses, serving ~100 Tb of data.

Saved Sessions usage

- The use of the Saved Sessions feature continues to increase. Figure B.4.3 shows the steady increase in the use of this feature since its introduction in 2007. Saved Session usage has risen to more than 1,200 users creating more than 3,000 new Saved Sessions per month, a good measure of active engagement with the Browser. A consistent 15% of session activity is from the European mirror, and an unknown amount of such activity takes place on private mirrors.

Figure B.4.3. New sessions created, and number of users creating them, since 2007 on the US + European + Asian servers combined (15% of the total takes place on the European mirror, a negligible amount on the Asian server). As the month-to-month variation in the creation of sessions is quite variable, it is presented here as a six-month moving average to make the trend more obvious.



Track Hub usage

- Track hub usage has increased steadily during the period. The ENCODE portal at Stanford generated more than 14,000 track hubs on the fly and displays them on the Genome Browser. Excluding those, 1,970 new track hubs were created per month (excluding our genome-euro and genome-asia mirror sites), up more than 10% above last year.

Custom Track usage

- Approximately 1.4 million custom track files were uploaded to the Genome Browser US and European servers during the period. Approximately 60,000 custom tracks have been saved to sessions for long-term use. The asian server is not receiving as much usage as anticipated with only 10,000 custom tracks uploaded for the period. There was a noticeable uptick after the workshops at the Korean Genome Organization (KOGO) meeting in February 2017, raising the number of tracks created in a six-month window by more than double (410 to 850). While these numbers are modest, they do offer some rare direct documentation on the effectiveness of workshops.
- Some 90,000 custom tracks were preserved in saved sessions.

Website and hardware infrastructure

- To assure optimal site performance, we periodically perform hardware and software maintenance on the servers that underlie the public website. This year we performed two major maintenance tasks:
 - upgraded the primary site file server (hgnfs1) to the latest operating system
 - reinstated backup procedure for hgnfs1 and the custom track host machine (customdb)
- During maintenance windows we make a priority of minimizing impact on our users. This year, the maintenance tasks required taking our primary public website briefly out of service.
- Consistent with past years, the UCSC Genome Browser sites continue to perform with high reliability and stability. In addition to the scheduled outages, there were two power outages on campus during peak usage hours (6AM - 5PM Pacific Time Mon–Fri) which were required by upgrades to high-voltage power switches. These resulted in 4 hours total peak downtime. During all outages, the genome-euro and genome-asia servers in Germany and Japan remained accessible (neither of them experienced any downtime this year).
- Our BLAT servers suffered from four instances of peak-time downtime during the year, resulting in a total of 16 hours of downtime, but only on specific genome assemblies. At no time were all the servers down, and among the human blat servers, only hg18 was down, for 2.5 hours.

Genome Browser Scientific Advisory Board

- The Scientific Advisory Board consists of prominent scientists in genetics and bioinformatics:
 - Aravinda Chakravarti, Johns Hopkins University
 - Ross Hardison, Penn State University
 - Tim Hubbard, King's College, London
 - Mary-Claire King, University of Washington
 - Robert Waterston, University of Washington Genome Sciences
 - Barbara Wold, California Institute of Technology
- The Board meets in Santa Cruz each year to offer guidance. Because of the uncertainty surrounding funding of the new 5-year grant proposal, no SAB meeting was held during the period covered by this report. The last meeting was Apr 2016. The next is expected in 2018.