ELENA ZHUKOVA/UC SANTA CRUZ PHOTO SERVICES

'WE COULDN'T LIVE WITHOUT IT': UCSC GENOME BROWSER AT 25

After a quarter of a century, the website remains an essential tool for navigating and understanding the genome. **By Michael Eisenstein**

Jim Kent knew a thing or two about coding. A veteran of the early days of personal computing, Kent developed software for Amiga, Atari and IBM computers throughout the 1980s, including pioneering tools for computer animation.

By 2000, Kent had made the leap to bioinformatics, developing the computational genome-navigation system that would become known as the University of California Santa Cruz (UCSC) Genome Browser. But his expectations were tinted by his industry experience. "I was really used to software lasting about three years," says Kent. His latest creation, he predicted, would fare no differently.

"It didn't happen," says geneticist David Haussler, scientific director of the UCSC Genomics Institute that runs the Genome Browser. This month marks the 25th anniversary of the browser, which has blossomed into an essential resource for biologists around the world. On any given day, more than 7,000 unique users engage with the site, which hosts genome data from humans, mice and more than 100 other animal species and viruses, and allows researchers to annotate and interrogate those sequences in countless ways.

Comparative genomicist Michael Hiller at Goethe University in Frankfurt, Germany, relies on the browser to investigate patterns of genome change across animal species to identify signatures of evolution. "Looking at the Genome Browser is a large part of what we and the other lab members do on a daily basis," he says. He has even incorporated the tool into a course he teaches. "That's part of the exercises," he says.

With a steady stream of updates and user-generated annotation tools, the Genome Browser shows no signs of slowing down. "The website took on a life of its own," says Max Haeussler, the project's principal investigator. But as with many US-based scientific projects in 2025, the real threat to its longevity comes from tightening of the purse strings, with massive cuts and restructuring at the US National Institutes of Health (NIH) putting this essential resource at risk.

Hitting the ground running

It began with the Human Genome Project. Haussler joined the effort in late 1999, with the goal of developing software that could identify and annotate genes in the newly assembled genome. But there was an issue. "It became clear to me very quickly that this problem of assembling the human genome from small pieces of sequenced DNA was extremely difficult, and they had no serious plan to do it," Haussler recalls. He turned to Kent, a skilled coder who was completing a PhD at UCSC focused on developing the Intronerator, a genome browser for mapping gene expression in the worm model species *Caenorhabditis elegans*.

In a frenetic programming crunch that spanned several weeks in 2000, Kent developed GigAssembler — 10,000 lines of code that Haussler describes as "the assembler that saved the day". It enabled the Human Genome Project to complete its first draft in time for its official unveiling on 26 June 2000.

Kent's initial interest was mostly pragmatic. "I wanted to get the human genome browser up just so that I could double-check my assembly more easily," he says. But when he and Haussler launched the first iteration of the UCSC Genome Browser on 7 July 2000, it was an instant success. The scientific community collectively downloaded roughly half a terabyte of data in the first 24 hours of the browser's operation, the team estimates — a staggering amount at the time.

The genome was then only about 90% complete and rife with gaps, but the world finally had a freely available window for surveying it. "Just being able to go to a browser view of early gene annotations or early cloning enterprises was like a magnet," says Barbara Wold, a molecular biologist at the California Institute of Technology in Pasadena. "It drew everybody into the community of users." Importantly, the browser also established objective 'mile markers' that researchers could use to map discoveries and variants in individual genomes.

Ian Holmes, a computational biologist at the University of California, Berkeley, also credits Kent and the UCSC team for normalizing the idea that genomic data should be free and easy to engage with and reuse. "The idea that everything should be available on the web and not desktop applications was still an extremely radical idea then," says Holmes. "They invented a tonne of file formats and just ways of storing data that we are widely using today."

Over the next few years, the Genome Browser updated its maps with the latest data from the Human Genome Project, as well as from other popular model species, including the mouse and rat. Kent also introduced a defining feature, based on his experience with the Intronerator: the ability for users to generate and share custom 'tracks' that selectively annotate key features of the genome. Tracks are interactive visual road maps, presented alongside the genome, that chart the location of important features such as regulatory elements and disease-associated gene variants. "We were way ahead of everybody with that feature, and I think that's really kind of why we ended up so much in the middle of the ecosystem," says Kent.

The effort gained early momentum thanks to buy-in from key international consortia such as the Encyclopedia of DNA Elements (ENCODE) project. ENCODE was focused on the functional motifs that populate the 98.5% of the human genome that does not contain protein-coding gene sequences. Haussler says that the Genome Browser's roster of custom tracks "grew dramatically" during the ENCODE team's work, allowing users to visualize the many promoters, enhancers and non-protein-coding RNAs that the consortium uncovered.

A durable resource

Still, Kent's initial gloomy prediction of the browser's life expectancy was not unreasonable — such narratives are typical in academia. "Imagine if Microsoft Office disappears every three years," says Haeussler. "This is the case for most bioinformatics-software packages." Graduate students and postdoctoral researchers move on, and platforms stagnate even as new data accumulate, until somebody else is compelled to develop a replacement.

And the data have steadily poured in. "There are no White House meetings any more when you publish a genome," says Haeussler, referring to the announcement of the first human genome draft. "Now we get one a day, or even ten a day." And it isn't only genomes, but also new layers of information that people want to map on top of them, such as full-length RNA transcripts and epigenetic signals that play a central part in gene regulation. At the same time, the tool's user base has grown steadily.

The UCSC Genome Browser's ability to survive and accommodate this growth is partly attributable to the early decision to build a team of dedicated programmers and engineers. Many of these were seasoned Silicon Valley veterans impressed by the browser's ambitions and scrappy origin story. "I kind of crashed [Kent's] PhD defence," says Angie Hinrichs, a former microchip designer who had minimal biology expertise but nevertheless volunteered for the Genome Browser team in early 2002. "And then, in May, when they were hiring, I just barely managed to squeak in over their requirements." She would be one of the first new hires on a team that swelled to roughly 30 people at its peak.

"Looking at the Genome Browser is a large part of what we do on a daily basis."

Access to affordable and powerful computing infrastructure hasn't hurt, either. "In the early days, it probably seemed like a lot to have 32 processors, and now our development machine has 192," says Hinrichs. The data themselves, as well as the code to run the browser, reside on petabyte-scale arrays that include numerous super-speedy, solid-state hard drives, she adds.

The Genome Browser's codebase has grown from just 10,000 lines to more than three million, Haussler says. But its fundamental architecture has not changed much in 25 years. Hinrichs says: "There have been many chefs in the kitchen, and it's gotten a little messy, but the core structure of it was so clean that it has actually worked and we can still develop on it. I credit Jim Kent for that."

From Kent's perspective, this durability is attributable to his focus on creating modular code that can be tweaked without bringing down the whole edifice, as well as a dedicated testing corps. "We had a quality-assurance team that was about half the size of the software-development team," he says. "I think we only had to roll back a previous version once."

The team has had to devise more efficient ways to handle data, however. In 2011, for example, it introduced the 'track hub' format as a solution to the increasingly large and complex custom tracks that external laboratories were developing for genome annotation. Track hubs comprise sets of relatively compact binary files that are hosted by the data creators themselves, rather than at UCSC. The data are streamed on demand, and "Jim's file formats are designed so you can jump between them," says Haeussler. The ability to rapidly and selectively access the relevant parts of the remotely hosted data minimizes the amount that needs to flow back and forth between the user and the Genome Browser. This maintains performance speed even as the number of tracks rises. "We have thousands of labs making them, but most of them we never see," says Haeussler.

And in 2023, the team launched the initiative GenArk (Genome Archive) to accommodate the unending influx of new genome sequences (H. Clawson *et al. Genome Biol.* **24**, 217; 2023). This system makes it straightforward for the Genome Browser to pluck newly uploaded genomes from the Assembly database hosted by the US National Center for Biotechnology Information (NCBI), allowing researchers to view and annotate those genomes. GenArk currently hosts more than 6,000 genomes, ranging from bacteria to primates.

Something for everyone

Much of the team's time is spent engaging with the users who shape the Genome Browser's content. The most common contact methods are a popular mailing list and forums, through which researchers can request technological support and propose ideas for tracks.

Track development tends to be highly collaborative. "It's not only that we work with the developers, but we also contact the data owners," says Anna Benet-Pagès, a bioinformatician at the Medical Genetics Centre in Munich, Germany, who collaborates regularly with the UCSC team. "Sometimes they might do things to even improve the data, or we talk to them on what would be the best visualization from their perspective for the data." This has resulted in a platform that is broadly useful to almost any area of biology that intersects with a genome.

For example, Benet-Pagès uses the Genome Browser to identify genetic variants that put

people at risk of cancer or other diseases. "With modern extensive sequencing of so many genomes, you can answer the question: here's a mutation that I think might be disease-causing, but what's its penetrance?" she says, referring to the likelihood that a disease-associated sequence change will result in a pathological condition. Benet-Pagès has assisted with track development for clinical genetics consortia such as the NIH ClinGen programme, which has assembled a database of disease-related mutations for more than 3,200 human genes. More sophisticated interpretation tools are available, but being a free public resource gives the Genome Browser an edge. "Genetics labs in many countries are not able to afford the prices of some software," says Benet-Pagès.

Hiller's lab relies heavily on the Genome Browser for its comparative genomics work. In 2023, his team developed a tool called TOGA, which is compatible with the browser and can identify and annotate evolutionarily related genes across species. This makes it possible to determine how gene loss, duplication or other changes contribute to distinctive traits (B. Kirilenko *et al. Science* **380**, eabn3107; 2023). He is currently focusing on bat genomics as part of the international Bat1K Project to sequence all living bat species, which aims to understand the distinctive and medically interesting traits of particular bat species, including resistance to viral disease, efficient metabolism and surprising longevity.

The browser is also powering pathogen surveillance. Early on in the COVID-19 pandemic, UCSC programmer Hiram Clawson developed a viewer for the SARS-CoV-2 virus sequence. As global sequencing efforts picked up steam, Hinrichs stepped in to expand and oversee the effort to compare and annotate past and emerging viral variants using data from major global repositories. She maintained "a tree of essentially all the SARS-CoV-2 genomes", she says. The browser has maps of the ancestry of more than 17 million viral isolates (see go.nature.com/4nc9vbq). Hinrichs recently left the Genome Browser team, and is working with UCSC evolutionary genomicist Russ Corbett-Detig to apply a similar approach to pathogens such as tuberculosis, the H5N1 avian influenza virus and monkeypox virus. "I sure hope there's not another pandemic any time soon, but we're ready if there is," she says.

Other genome-visualization tools benefit from this open-data ecosystem. For example, Holmes and colleagues have developed a software tool called JBrowse that allows users to rapidly navigate and annotate genomes that are relevant to their research. Users can integrate their views with the tracks that are available through the UCSC Track Hub Registry. "It's a very enlightened thing that they're doing," says Holmes of the UCSC team. "All that data-integration work that they've done is available to users of JBrowse and other browsers as well, so we can just connect to that."

NHGRI/NIH

Senator Tom Harkin (Democrat) helped to secure funding for the Human Genome Project, which was able to assemble the sequence using code from the UCSC Genome Browser team.

Opportunities and threats

After a quarter of a century, many frontiers remain to be charted, and the Genome Browser is well positioned to supply the maps.

For example, as a former member of ENCODE, Wold is deeply interested in characterizing the regulatory landscape that guides human gene expression. She cites the work of the Human Pangenome Reference Consortium, which is building a more representative assembly of the human genome that encompasses the variability of hundreds of people worldwide. "The effort to capture the varied genomes is now a frontier for browsing," she says. "How are we going to be able to capture that at will, in a way that a user can unwrap?"

Haussler sees exciting opportunities to bolster the Genome Browser with insights from analyses of biological data that use artificial intelligence. He is in touch with members of the Virtual Cell project, an effort led by the Chan Zuckerberg Initiative to generate sophisticated models of cellular biology by applying deep learning to vast quantities of experimental data. Such a simulation, Haussler says, could generate well-informed predictions about how changes in sequence or structure inform genome function, and reveal their consequences for human health. The model will be "incomprehensibly richer than any model we've ever applied," he says. "Then we can finally go back and start asking the questions that everyone asked when we first did the Human Genome Project."

Funding remains a problem, however. The Genome Browser initiative has historically received up to US$4 million per year in funding from the US National Human Genome Research Institute. That budget has been strained in recent years by rising costs for staff and equipment, and the current team includes just eight dedicated members. With NIH funding now being gutted and many of its constituent institutes at risk, the financial future of the Genome Browser project is unclear.

This uncertainty is frustrating and unsettling to Haeussler. "It's one of the most cost-efficient projects of the NIH, if you calculate it per user," he says. He is now working with other NIH bodies to secure funding, and weighing seeking out international support given the Genome Browser's vast user base in Europe and Asia. Other organizations host complementary tools that offer similar genome-analysis capabilities, including the European Bioinformatics Institute's Ensembl browser and the NCBI Genome Data Viewer. But some labs have literally built their analytical workflows around the UCSC site. "We couldn't live without the browser," says Hiller. "If the browser goes away, I can shut down the lab."

Michael Eisenstein is a science writer in Philadelphia, Pennsylvania.