

This grant has supported the UCSC Genome Browser website and associated tools and databases for five years. The UCSC Genome Browser is a free web-based tool used by over 100,000 biomedical scientists. The Genome Browser hosts a wide range of data useful for understanding the DNA of humans and other species. The Browser was first developed in a simple form in conjunction with the Human Genome Project in 2000, and has been extended and enriched continuously ever since.

During the five years of this grant UCSC extended the browser and associated tools and websites in many ways, keeping up with a biomedical field that is exploding in terms of capabilities and data. The data we now show includes variation in DNA observed both commonly and rarely in humans in general, in patients afflicted with diseases in particular, and in animals used in research, agriculture, and evolutionary studies. We've refined the definitions of what parts of the total DNA correspond to what genes, and what genes are active in various human organs, tissues, and cells. These genes include the classic type of gene where DNA is transcribed into RNA and then translated into functional protein molecules, as well as genes that are functional directly at the RNA level. We also host and display data on DNA that regulates gene use; that is responsible for turning genes on and off where they are needed.

The data we display and analyze comes from a wide variety of sources. A substantial fraction is from large consortium projects such as the 1000 Genomes project for common human variation, the GTEx project to define tissue specific gene expression and how it varies between individuals, and the ENCODE project which seeks to define gene regulatory regions. A substantial fraction of our data comes from the research of individual labs and scientific publications. We developed a method that mines DNA sequence from scientific publications, and provides quick links from the sequence in the Genome Browser to the associated publication that is especially useful for finding relevant work from smaller groups. We also have developed mechanisms so that individual scientists can view their own, potentially unpublished and confidential, work in the context of the Genome Browser.

We've extended the software that powers the Genome Browser as well. The web site became more interactive as advances in web technology allowed us to push past click/wait/read/click user interfaces to a richer user experience including drop-down menus, auto-suggested search terms, right click menus, and drag select. We've scaled up and optimized the site to handle bigger data sets and more users without lagging. We've added security and privacy layers so that users need not fear that their unpublished or confidential data will leak out. We've developed new tools aimed at addressing a central question of the field – which DNA variations present in a particular person are likely to have an effect on their health.

In all, this grant has supported the work of an active software engineering and data coordination group at UC Santa Cruz in its efforts to provide useful tools for humans to understand the human genome in all of its complexity and diversity.