

## [UC Santa Cruz Genomics Institute](#) Portfolio: An Overview of Our Research Projects

Led by scientific director and UC Santa Cruz professor of biomolecular engineering David Haussler, the UC Santa Cruz Genomics Institute provides world leadership in data analysis, precision medicine, and technology platforms. In fact, UC Santa Cruz is the premier institution on a global level for handling big data in genomics research. We are unmatched in storing, cataloging, assembling, validating, and analyzing huge volumes of genomic data. Our mission is to use genomics to positively impact health, nature, and society.

Approximately half of all diseases have a substantial genomic component, including contributions from millions of individually rare but collectively common genetic variations. Genomic information is essentially massive quantities of data. Continual innovation in information technology enables faster, more efficient analyses of huge data sets. With this analysis, will come wide deployment of precision diagnosis and treatment based on genomic information. Only by studying very large numbers of individual genomes will scientists have the statistical power to discover and understand comprehensively the genomic contribution to disease.

And human health is just the beginning. Browse our project portfolio below to learn about research into biodiversity and efforts to ensure society uses the power of genomics thoughtfully and responsibly. You may also visit our research group websites: [UCSC Genome Browser](#), [Genome 10K](#), [Global Alliance for Genomics and Health](#), [Computational Genomics Laboratory\(CGL\)](#), [Treehouse Childhood Cancer Initiative](#) and [BRCA Exchange](#).

### [The Treehouse Childhood Cancer Initiative](#)

The mission of the UC Santa Cruz Genomics Institute's [Treehouse Childhood Cancer Initiative](#) is to change the story for childhood cancer patients by leveraging genomic data and computational approaches that could identify less toxic and more effective treatments. Treehouse analyzes a child's cancer data against both childhood and adult patient cohorts across all types of cancer. This "pan-cancer" analysis of adult and pediatric tumors may identify situations where an adult drug is predicted to work on a subset of pediatric patients.

We are able to do this work at UC Santa Cruz because we can compare individual pediatric tumors against the vast database of 11,000+ tumors available at the UC Santa Cruz Genomics Institute, which is internationally known for its commitment to sharing genomic data worldwide.

Initially funded by the California Initiative to Advance Precision Medicine, Treehouse is now funded by St. Baldrick's Foundation, with additional gifts from local pediatric cancer advocacy organizations such as Unravel and Team G.



### **Computational Genomics Lab**

Dedicated to furthering open-platform genomic science and technology, the **Computational Genomics Lab** is leading efforts at the UC Santa Cruz Genomics Institute to connect isolated genomic data silos and forge platform standards; to build more complete and accurate genomic information; to build a human reference genome that contains all common variations; to discover and share portable, reproducible and scalable software workflows that will allow bioinformatics scientists to work better, together; and to analyze massive genomic datasets using the cloud. The goal is to provide transformative resources for global research communities who rely on genomic data.

### **Global Alliance for Genomics and Health (GA4GH)**

This non-profit global alliance co-founded by David Haussler was formed to help accelerate the potential of genomic medicine to advance human health. It brings together over 400 leading institutions working in healthcare, research, disease advocacy, life science, and information technology. Member organizations agree to work together to create a common framework of harmonized approaches to enable the responsible, voluntary, and secure sharing of genomic and clinical data. The GA4GH's Data Working Group, which David Haussler co-chairs, concentrates on developing standards for how to represent, store and analyze genomic data. These activities include working with academic and industry leaders to develop approaches that allow different programmers and computer systems to work together seamlessly.

### **BRCA Network**

The BRCA network is a global open source resource governed by the Global Alliance for Genomics and Health that is based at the UC Santa Cruz Genomics Institute. The network hosts clinical breast-cancer data contributed by geneticists from several countries, and report on more variants than any other public repository. Through this project we have developed [brcaexchange.org](http://brcaexchange.org), the world's largest public, open repository of information on BRCA1 and BRCA2 genetic variations and their implications for cancer risk.

### **Human Cell Atlas (HCA)**

The UC Santa Cruz Genomics Institute collaborates with partners in the United States and Europe on the Chan Zuckerberg Initiative [Human Cell Atlas](#) (HCA) Data Sharing Platform. The HCA is designed to be an invaluable public reference tool that will revolutionize our understanding of the human body by using genomics to map and characterize every human cell type.

### **Center for Big Data in Translational Genomics**

This multinational collaboration between academia and industry creates data models and analysis tools to analyze massive data sets of genomic information. This is the infrastructure for the broad application of genomics in biomedicine. Such tools are used to analyze data from thousands of individuals to uncover the contribution of gene variants to disease, with an initial focus on cancer. This knowledge is instrumental in the development of precision diagnostic and treatment methods.

### [Dockstore](#)

The UC Santa Cruz Genomics Institute has provided global leadership in developing tools that work across clouds to provide large-scale analysis of cancer data using our best-practices tool and workflow repository, [Dockstore](#). Dockstore offers scientists a platform to package up their analyses and send them to data in order to analyze in place, avoiding time-consuming and costly downloads of very large datasets. Our goal is to create a Dockstore-based, federated network of globally distributed, searchable repositories. We believe this will make it much easier for researchers to find, share and use biomedical software, which will be a significant benefit to the broader biomedical research community.

### [GENCODE](#)

[GENCODE](#), a subproject of the ENCYclopedia of DNA Elements (ENCODE) project, aims to build an encyclopedia of genes and gene variants. GENCODE will build this reference tool by identifying all gene features and functions in both the human and mouse genomes. It will use a combination of computational analysis, manual annotation, and experimental validation. It will annotate all evidence-based gene features (gene functions that can be validated through experiments against a reference gene) in the entire human genome at a high accuracy.

### [Stem-Cell Hub](#)

The UC Santa Cruz Genomics Institute [Stem-Cell Hub](#) houses all California Institute of Regenerative Medicine (CIRM) genomic data, providing an invaluable open resource to the biomedical research community. It supports stem cell clinical trials for diseases such as Parkinson's Disease and Cardiomyopathy, among others.

### [NIH/NHLBI TOPMed Consortium](#)

The UC Santa Cruz Genomics Institute has led the global effort to capture the natural diversity of the human genome sequence using a mathematical structure called genome graphs. Our goal is to make genomes more accurate, more complete and no longer biased toward a particular ethnic group that contributed to the single reference human genome we use today. We are using this approach to provide unbiased analysis of all genomic variations found by the [NIH/NHLBI TOPMed Consortium](#) with a particular focus on hemoglobin disorders that disproportionately affect certain genetic subpopulations.

### [Human Genome Variation Map](#)

The Human Genome Variation Map (HGVM) takes the approaches developed in the TOPMed Consortium and applies them to an enormously ambitious goal of creating the first standard and comprehensive taxonomy for human variation. This map and the process to create it will do no less than transform genetics. Instead of describing genetic variations with respect to a changing, linear coordinate system (the current reference genome), it will add this missing variation to the reference, resulting in a structure that can be described as a mathematical graph: a genome graph.



### [UCSC Xena](#)

[UCSC Xena](#) is a bioinformatics tool used to visualize functional genomics data from multiple sources simultaneously, including both public and private data. The Xena system consists of a set of federated data hubs and the Xena browser, which integrates across hubs, providing one location to analyze and visualize all data. This allows researchers to combine new or preliminary results from their laptops or internal servers, or even data from a new research publication, securely with vetted data from the public sphere.

### [UCSC Genome Browser](#)

UC Santa Cruz is renowned for its strength in bioinformatics -- the investigation and analysis of biological data using computational, mathematical, and statistical approaches. The flagship product of this expertise is the UCSC Genome Browser.

The browser serves as an interactive web-based "microscope," allowing researchers to view all 23 chromosomes of the human genome at any scale, from a complete chromosome to an individual nucleotide. The browser integrates the work of countless scientists in laboratories worldwide, including work generated at UC Santa Cruz, in an interactive, graphical display.

Far from simply displaying the genetic code, the UCSC Genome Browser brings the code to life by aligning relevant areas with experimental and computational data and images. It also links to international databases, giving researchers instant access to deeper information about the genome. An experienced user can form a hypothesis and verify it in minutes using this tool. The UCSC Genome Browser receives over a million hits daily. As a platform, it has multiple potential uses that can improve diagnosis, support prevention, and lead to cures for disease.

### [Genome 10K](#)

The Genome 10K project aims to assemble a genomic zoo — a collection of DNA sequences representing the genomes of 10,000 vertebrate species, approximately one for every vertebrate genus. The Genome 10K project was co-founded a decade ago by the Genomics Institute's Scientific Director, David Haussler, along with Steve O'Brien from the National Cancer Institute and Oliver Ryder from the Zoological Society of San Diego. Capturing the genetic diversity of vertebrate species would create an unprecedented resource for the life sciences and for worldwide conservation efforts. Genome 10K's global community of scientists representing major zoos, museums, research centers, and universities around the world is well on its way to completing its collection of tissue specimens and genome sequences.

### [Nanopore Sequencing](#)

UC Santa Cruz researchers invented the idea of [nanopore sequencing](#). They developed [nanopore detectors](#) -- single-cell, long-read genetic sequencing technologies -- which can be compared to the impact of the electron microscope to science in the 20th century. These nanopore instruments are built

around a “nano-scale” opening in a thin membrane, for the study of microscopic living material such as DNA. These pores or openings are just big enough to fit a single strand of DNA.

### [Nanopipette Technology](#)

UC Santa Cruz researchers have also developed a [nanopipette technology](#) to sample and analyze living cells without harming them. This amazing innovation allows researchers to perform longitudinal genomic studies on single cells in vivo, which, for example, could pave the way to understand neurodegeneration occurs, and how cancer cells develop resistance to chemotherapy drugs.

### [Paleogenomics Lab](#)

The Paleogenomics Lab is a joint venture developing new experimental and computational approaches to recover DNA preserved within the remains of organisms that lived tens to hundreds of thousands of years ago. Their research focuses on a wide range of evolutionary and ecological questions, mostly involving the application of genomics techniques to better understand how species and populations evolve through time. Their interests are broad, spanning human evolution, genome assembly and analysis, pathogen evolution, population genetics, conservation genomics, and the genomic consequences of long-term environmental change. They are also developing approaches to use genome editing technologies to resurrect extinct traits, and possibly extinct species, as a potential new tool for biodiversity conservation.

### [Science & Justice Research Center](#)

The Science and Justice Research Center (SJRC) works to infuse science with commitments to justice. It believes science is not just about what happens at the lab bench; it shapes the nature of our present and future. SJRC informs and trains a new generation of leaders who can create science and technology that are responsive to diverse needs and tackle complex, pressing problems, including race, health and social justice, conservation of ecologies and ways of life. By bringing together diverse international leaders, it aims to shape big biodata and precision medicine’s science and justice agenda and broaden the public discussion to address fundamental questions about the right and just constitution of care, trust, and knowledge in an age of biomedical data.

## Funding

The UC Santa Cruz Genomics Institute and the [Hausler Lab](#) receive generous funding for our research and staff from key private and public foundations and institutes. This work has been supported to date by more than \$150 million in federal, state and private extramural research funding. For a complete list of our affiliations and funding institutions, please visit our [website](#).