UC Santa Cruz Genomics Institute Backgrounder: History & Research Portfolio

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.

History of the Genomics Institute at UC Santa Cruz

The modern genomic age started more than 30 years ago when a group of eminent biologists gathered in Santa Cruz to propose a massive and historic project to determine the complete DNA sequence of the human genome — our genetic blueprint. By 1999, research teams at government institutions, universities and corporations around the world were vying to be first to sequence the human genome. Despite robust cooperation within a non-profit International Human Genome Sequencing Consortium, a real concern grew that a corporation might obtain the human genome sequence before any others and patent key parts of it, restricting access to the holders of the patent and those who could afford licenses or subscriptions, thus severely limiting research.

It was then that UCSC's David Haussler joined the international consortium and enlisted molecular biology graduate student Jim Kent to make sure this didn't happen. They made history when they assembled the DNA sequencing data from the international consortium just days before a competing corporate team assembled theirs and posted the first human genome sequence on the internet on July 7, 2000. That put the human genome in the public domain. ensuring our genetic code would always be freely available for all of humanity to share.

Organized as the Genomics Institute since 2014, UCSC's genomics investigators have built a research powerhouse in sequencing, storing, cataloging, assembling, validating, and analyzing huge volumes of genomic data in their mission to use genomics to positively impact health, nature, and society. Without a medical school and hospital, UCSC has become identified as a neutral, trusted genomics research partner to medical institutions and universities around the world, contributing databases and research, and improving human health.

Comprised of 150 faculty, affiliates, staff and student researchers, the UC Santa Cruz Genomics Institute draws strength from its unique nexus of diverse perspectives on health, nature, technology and social justice. Fuelled by a vision of using the power of genomics collaboratively, openly and ethically for the benefit of both the individual and the planet, the Institute is engaged in the practical work required to make this vision a reality.

Haussler Lab: Genetic Roots of Brain Function, Disorders

David Haussler's prestigious designation since 2000 as an HHMI investigator has provided the Haussler Lab freedom to conduct groundbreaking research. The Haussler lab investigates the evolution and function of non-protein coding regions of the human genome. A major focus is on identifying the genetic elements that have played a role in the development and evolution of the human brain. Its innovations range from "mini-brains" — tiny collections of cells complex enough to mimic brain function — to new capabilities for the vast information infrastructure required for large-scale medical genetics. Its goal is to create a novel medical resource that can be used to uncover the role of specific parts of our genome in both normal brain function and in a range of brain disorders, including autism, schizophrenia, developmental delay, bipolar disorder, and many others.

Data Sharing to Advance Medical Research

Leading the charge towards international data sharing, Haussler co-founded the **Global Alliance for Genomics and Health (GA4GH)** in 2013. Since then, this nonprofit has worked to improve the potential of genomic medicine to advance human health. The Alliance brings together over 500 leading institutions working in healthcare, research, disease advocacy, life science, and information technology, collaborating to create a common framework of harmonized approaches designed to enable the responsible, voluntary, and secure sharing of genomic and clinical data.

In research across all cancers, **Josh Stuart's lab** is leading the **Cancer Genome Atlas'** groundbreaking **Pan-Cancer Initiative**, analyzing digital data to reveal similarities between tumors in different parts of the body that were not previously apparent. This research is upending a world where tumors have been classified and treated based on where in the body they originated, rather than by considering their molecular origins.

Angela Brooks' cancer research has focused on alternative splicing and the identification of mutations that lead to meaningful changes in gene expression. Using powerful new methods for identifying mutations, Brooks has investigated proteins involved in splicing and human cancers and successfully identified the cancerous effects of mutations in two genes: U2AF1 in lung adenocarcinoma and acute myeloid leukemia2 and SF3B1 in chronic lymphocytic leukemia.

The Genomics Institute co-founded the **BRCA Exchange**, a global open source resource governed by the **Global Alliance for Genomics and Health (GA4GH)**. This network hosts clinical breast-cancer data contributed by geneticists from several countries, and has become the world's largest public, open repository of information on BRCA1 and BRCA2 genetic variations and their implications for cancer risk.

In pediatric cancer, **The Treehouse Childhood Cancer Initiative** is changing the story for childhood cancer patients by leveraging genomic data and computational approaches to identify less toxic or

more effective treatments. Treehouse analyzes a child's cancer data against both childhood and adult patient cohorts across all types of cancer. It can compare individual pediatric tumors against a vast database of 11,000+ tumors. This "pan-cancer" analysis of adult and pediatric tumors may predict situations in which an adult drug might work on a subset of pediatric patients.

In cooperation with our sister organization the Institute for the Biology of Stem Cells, led by faculty members Camilla Forsberg and Lindsay Hinck, The Genomic Institute also coordinates and manages the **Stem Cell Hub**, a project of the California Institute of Regenerative Medicine (CIRM). CIRM's goal for the Stem Cell Hub is to apply genomics and bioinformatics approaches to stem cell research, providing an invaluable open resource to the biomedical research community and supporting clinical trials for diseases such as Parkinson's disease and cardiomyopathy. Among those contributing to this project are faculty members **Daniel Kim**, **Nader Pourmand** and **Josh Stuart** from UCSC's Biomolecular Engineering department; and **Jeremy Sanford** from UCSC's Molecular, Cell, & Developmental Biology department.

Nature, Prehistoric Creatures and Ethical Research

The **Genome 10K** project, co-founded a decade ago by Haussler, 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. Genome 10K is capturing the genetic diversity of vertebrate species to 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.

Genomics Institute researchers led by faculty members **Beth Shapiro**, **Ed Green** and **Lars Ferhen-Schmitz** are also making enormous and exciting progress in paleogenetics, including creating new ways to recover DNA preserved within the remains of organisms that lived hundreds of thousands of years ago. They are 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 innovations are helping to revise long-held beliefs about how life came to be, helping to put dates on culturally (and climatically) significant events such as the arrival of bison in North America (NY Times 3/13/17).

The **Science & Justice Research Center** (SJRC) at UCSC, lead by founding director and UCSC Sociology professor Jenny Reardon, works to infuse genomics -- and all 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.

Bioinformatics Research, Platforms and Tools to Mine Data

The newer genomics tools and platforms developed at the Genomics Institute have contributed to a growing global, open-source infrastructure and are just the latest developments in a longer story of genomics leadership. Thanks in great part to its leadership role in the Human Genome Project, UC Santa Cruz has become 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**, led by principal investigator **Jim Kent**, which now receives over a million hits daily. As a platform, it alone has multiple potential uses that can improve diagnosis, support prevention, and lead to cures for disease.

Dedicated to furthering open-platform genomic science and technology, the goal of the **Computational Genomics Lab** of the Genomics Institute is to provide transformative resources for global research communities who rely on genomic data. Its **Computational Genomics Platform** (CGP) group is focused on developing computational platforms based on the research of CGL. The CGP's portfolio includes the **Dockstore** workflow repository, Toil workflow engine, Redwood cloud storage system, and the Boardwalk visualization portal. In affiliation with the GA4GH, Dockstore's ultimate goal is to create a federated network of globally distributed, searchable data repositories. Using Dockstore, global partners are already able to run high-level genomics pipelines on-site, overcoming legal barriers, simplifying data sharing, and reducing cost.

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.

Together with the Broad Institute of MIT and Harvard and the University of Chicago, the UC Santa Cruz Genomics Institute is forming the Commons Alliance to design and develop a cloud-based data storage and computation platform dedicated to biomedical research. Development at UC Santa Cruz of the **NIH-funded Data Commons platform** will be directed by **Benedict Paten**. The platform will be modular, community-driven, open and standards-based, consistent with standards developed by coalitions such as the Global Alliance for Genomics and Health (GA4GH). It will be designed to accommodate a diversity of data types, including clinical and imaging data.

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 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.

Among other exciting CGL and CGP projects are the **Center for Big Data in Translational Genomics;** coordinated as one of the NIH's Big Data to Knowledge BD2K centers, this is a multinational collaboration between academia and industry creating data models and analysis tools to analyze massive data sets of genomic information. **GENCODE** is a subproject of the ENCylcopedia of DNA Elements (ENCODE) project, aiming to build an encyclopedia of genes and gene variants. The group is also working on an 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. Its work on the **Human Genome Variation Map** (HGVM) is using the approaches developed in the TOPMed Consortium and applying 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.

Nano Technology

In the technology realm, UC Santa Cruz researchers led by professors David Deamer and **Mark Akeson** invented the idea of **nanopore sequencing**, which now allows quick and inexpensive sequencing in the field. This technology has been adapted and extended by others like UCSC's **Christopher Vollmers** and is poised to become the *dominant genetic sequencing technology on the planet* within the next few years. UCSC professor **Nader Pourmand** also invented **nanopipette technology**, which can help researchers study microscopic living material. This amazing innovation allows researchers to perform **single-cell surgery**, allowing researchers to study living cells without harming them and to perform longitudinal genomic studies on single living cells. That could pave the way to understanding how neurodegeneration occurs, and how cancer cells develop resistance to chemotherapy drugs.

Technology Incubation

UC Santa Cruz Genomics Institute leadership is also involved in forming a campus-affiliated technology incubator, called **Startup Sandbox**, which aims to foster more technology startups in Santa Cruz like local spin-offs Two Pore Guys and Dovetail Genomics. The incubator is also supported by collaborators at the UCSF Mission Bay-based partnership QB3 and hometown allies like Santa Cruz Works. Startup Sandbox is just beginning to install lab benches and post "for lease" signs at its new home on Santa Cruz's Westside. It's a modest but growing infrastructure designed to inspire entrepreneurs to forgo a commute to Silicon Valley just over the hill in order to live, work and play by the beach.

Our UCSC Home: The Jack Baskin School of Engineering

Home to the UC Santa Cruz Genomics Institute, the Baskin School of Engineering at UC Santa Cruz offers unique opportunities for education, research and training. Faculty and students seek new approaches to critical 21st century challenges within the domains of data science, genomics,

bioinformatics, biotechnology, statistical modeling, high performance computing, sustainability engineering, human-centered design, communications, optoelectronics and photonics, networking and technology management. By leveraging novel tools that emerge from changing technologies, we have pioneered new engineering approaches and disciplines, examples of which include biomolecular engineering, computational media, and technology and information management.

The Genomics Institute unites affiliates in UCSC's Division of Physical and Biological Sciences (PBSci) and UCSC's Division of Social Sciences. For additional details on the Genomics Institute's affiliations and funding institutions, visit genomics.ucsc.edu/affiliations.

Funding

The UC Santa Cruz Genomics Institute and the Haussler 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.

Additional Resources

For a complete view on the Genomics Institute's innovative path to life science and medical research, visit its research group websites mentioned here: The UCSC Genome Browser, Computational Genomics Laboratory (CGL), the Treehouse Childhood Cancer Initiative, the Haussler Lab, the Stuart Lab, the Brooks Lab, the Nanopore Group, the Biosensor & Biotechnology Group, the Daniel Kim Lab, the Vollmers Lab, the UCSC Paleogenomics Lab, the UCSC Human Paleogenomics Lab and UCSC's Science & Justice Research Center.