UCSC Genome Browser: visualization for integrated analysis of diverse genomic data

Angie S Hinrichs¹, Ann S Zweig¹, Robert M Kuhn¹, Donna Karolchik¹, UCSC Genome Bioinformatics Group¹, David Haussler¹,², W James Kent¹

¹University of California Santa Cruz (UCSC), Center for Biomolecular Science and Engineering, School of Engineering, Santa Cruz, CA, 95064, ²Howard Hughes Medical Institute, UCSC, Santa Cruz, CA, 95064

The University of California at Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/) provides a searchable, interactive graphical display of public genome data and annotations such as phenotype and disease associations, gene predictions, transcription, expression, regulation, comparative genomics, and variation data. As the Data Coordination Center for the Encyclopedia of DNA Elements (ENCODE) project, UCSC hosts data from over 1500 experiments, comprising dozens of assays in over 250 human cell and tissue types, including maps of chromatin features and DNA binding sites for over 200 factors and marks as of May 2011 with significant expansion anticipated. In addition to hosting many high-quality datasets, one of the Genome Browser's main strengths is the ability to display users' data in a variety of data formats as "custom tracks" alongside the hosted data tracks.

Datasets of particular interest to the medical genetics community include the Online Mendelian Inheritance in Man (OMIM), the DECIPHER database of submicroscopic chromosomal imbalance, Genetic Association Database (GAD), the National Human Genome Research Institute's Catalog of Genome-Wide Association Studies (GWAS) SNPs, dbSNP, SNP Arrays, HapMap SNPs and the Database of Genomic Variants (DGV).

The UCSC Genome Browser continues to evolve to meet the dynamic needs of its users. Recent large sequencing projects such as the 1000 Genomes Project have produced terabytes of data and analyses along with flexible data formats tailored to the new data, such as Binary Sequence Alignment/Map (BAM) for alignments of next-gen sequencing reads to a reference genome assembly and Variant Call Format (VCF) for SNPs, indels, CNVs and structural rearrangements. The Genome Browser now displays BAM and VCF with specialized graphics and configuration options.