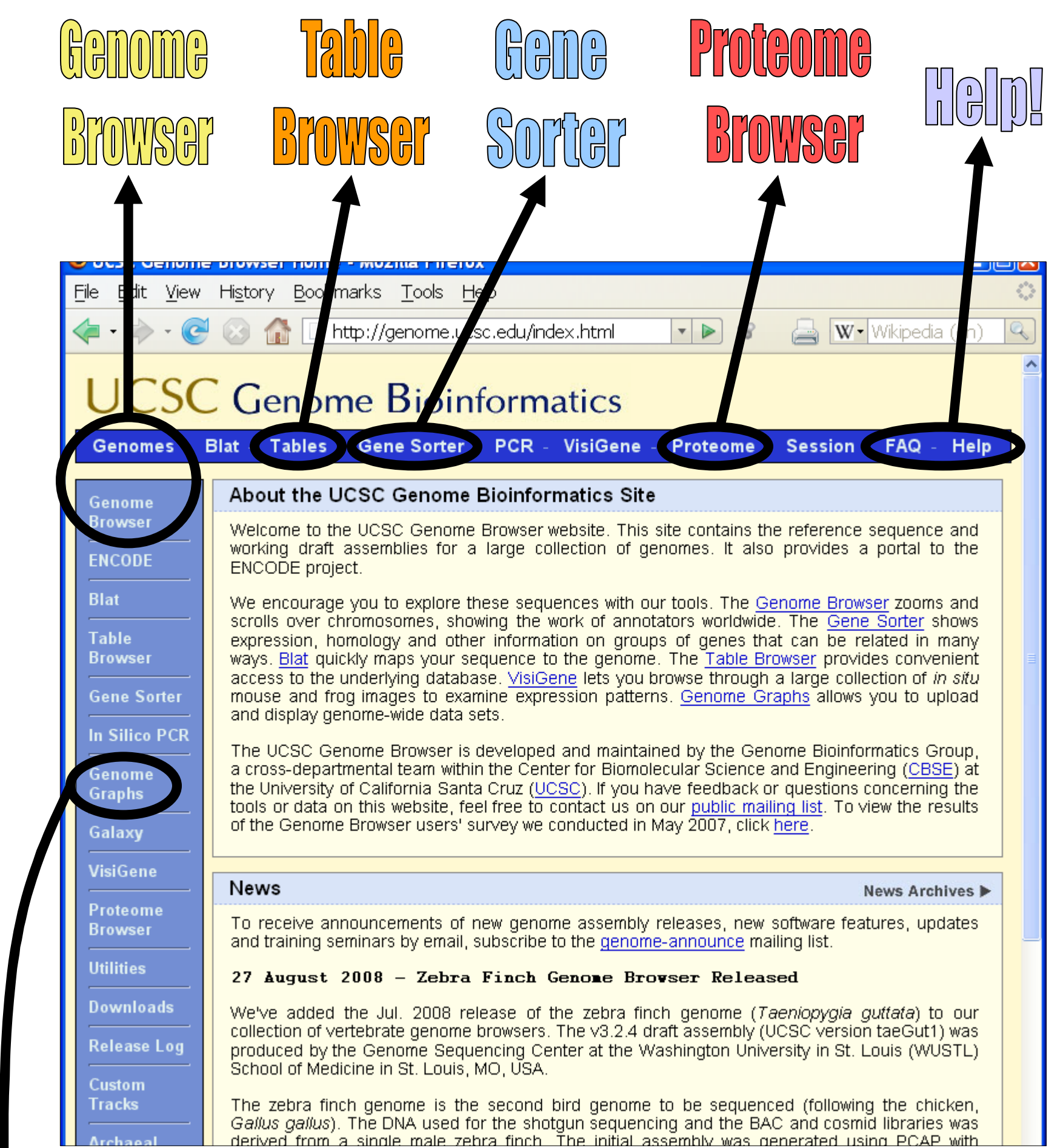


A. S. Hinrichs, R. M. Kuhn, D. Karolchik, A. S. Zweig, K. E. Smith, K. R. Rosenbloom, B. Rhead, B. J. Raney, A. Pohl, M. Pheasant, L. Meyer, F. Hsu, R. A. Harte, B. Giardine², P. Fujita, T. R. Dreszer, M. Diekhans, H. Clawson, G. P. Barber, D. Haussler¹ and W. J. Kent
 Center for Biomolecular Science and Engineering and ¹Howard Hughes Medical Institute, University of California Santa Cruz (UCSC); ²Center for Comparative Genomics and Bioinformatics, Huck Institutes of the Life Sciences, Pennsylvania State University

http://genome.ucsc.edu/ -- Visualization and Querying of Genomic Data

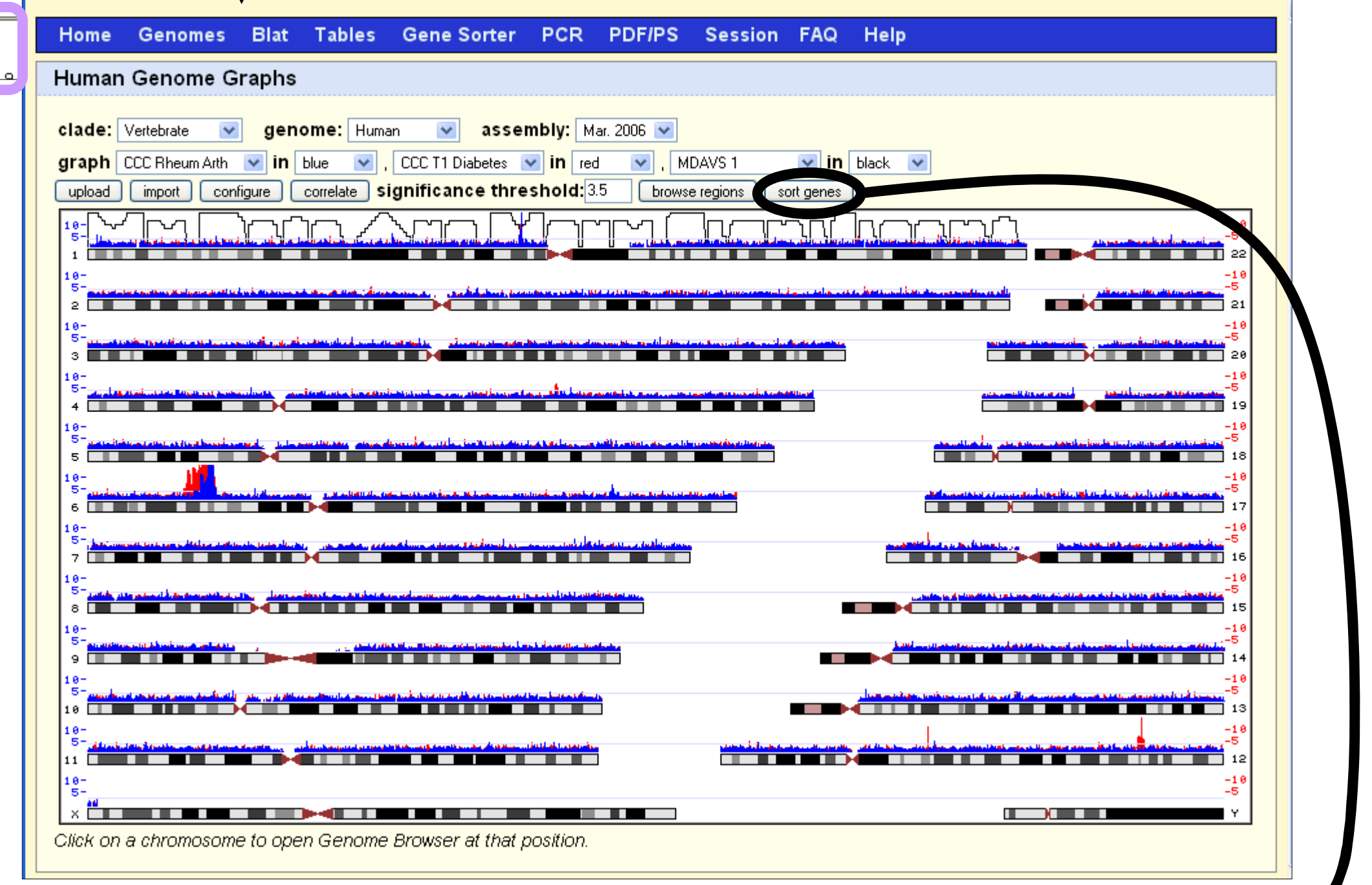
An effective platform for visualization and querying is invaluable when interpreting the mountains of data produced by high-throughput methods, especially with the advent of new sequencing technologies. The UCSC Genome Browser and other tools available at <http://genome.ucsc.edu/> serve this need by providing rich visualization, simple and complex querying tools, many hosted public datasets and annotations, and the ability to display and query uploaded data. Data mapped to reference genome coordinates can be displayed at any scale from base level to whole genome. The Genome Browser's graphical representations include simple box-and-line, histogram plots for numeric quantities, mountain plots for linkage disequilibrium, color-scale expression levels, base/codon labels at base level scale, and more. The display is highly configurable, and clicking on any item leads to more detailed information. Query support includes name, accession or keyword search (Genome Browser), sequence similarity search (Blat/isPcr), gene similarity by properties such as expression profile and GO terms (Gene Sorter), and operations on the underlying data: filtering, intersection, correlation (Table Browser). Hosted data types include SNPs, structural variation, repetitive elements, multi-species alignments and conservation scores, gene annotations, microarray probe mappings and expression levels, regulatory elements and targets, GenBank cDNA, genome-wide association results and all data produced by the ENCODE project. We are expanding our collection of human variation data and are preparing for the challenge of displaying the results of the 1000 Genomes project. Researchers can upload data in several straightforward flat file formats for viewing and comparison with the hosted data. In addition to the help pages on the site, a user-editable website (<http://genomewiki.ucsc.edu/>) provides additional documentation and we devote significant resources to answering questions sent to the publicly archived email list genome@soe.ucsc.edu.



Genome Graphs

Easily upload your own data!

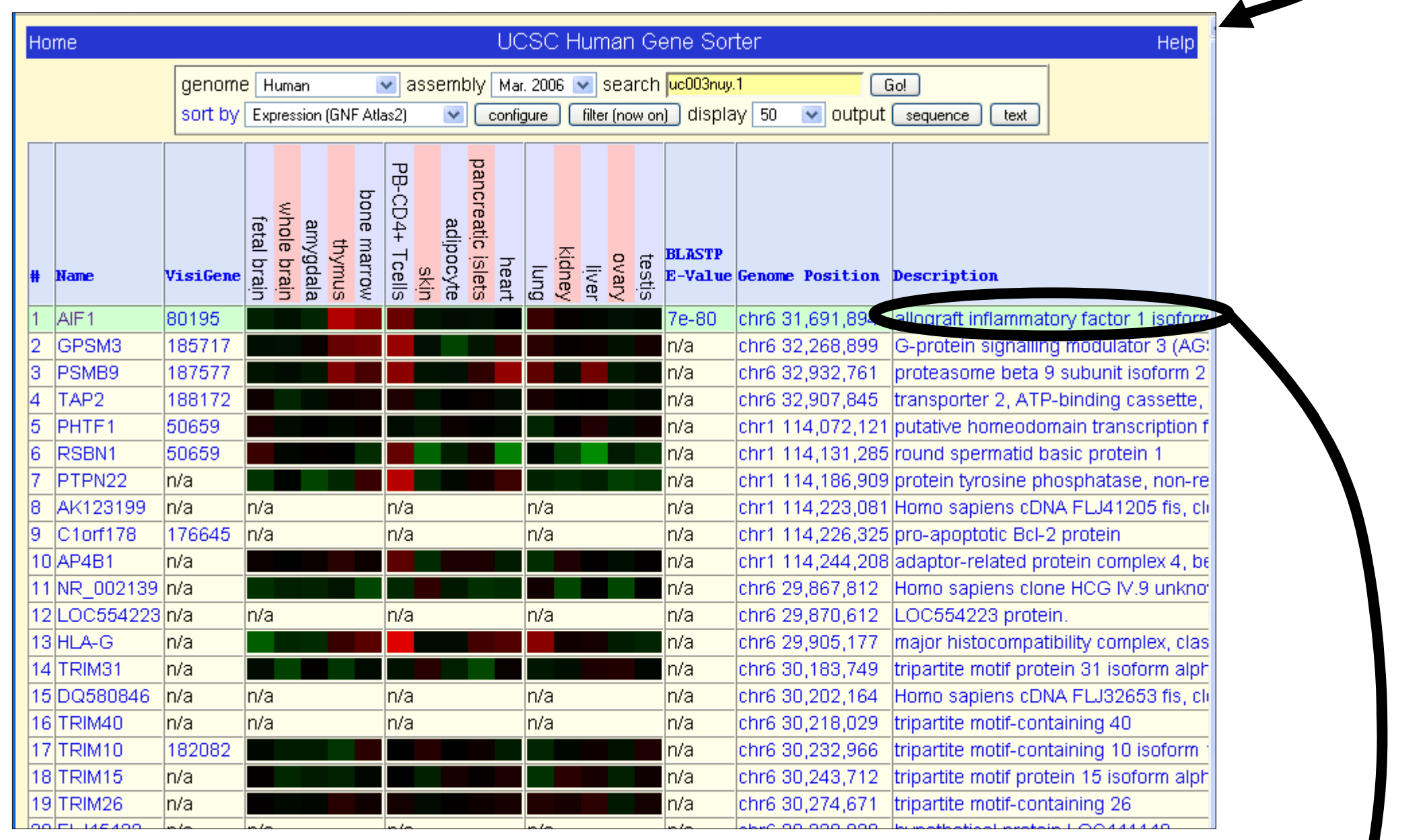
This tool plots numerical scores such as Genome-Wide Association Study results* against genomic position for all chromosomes in the genome. In addition to hosted data from several studies, scores can be uploaded in a very simple text format for viewing. Genome Browser tracks can also be imported for viewing. Up to four graphs per line can be superimposed. Clicking on a chromosome band leads to that band's region in the Genome Browser; clicking on a data point leads to the surrounding 1,000,000 base pair region. The significance threshold can be adjusted to select genes with sufficiently high scores, to view as hyperlinks to their regions in the Genome Browser, or to view that set of genes in the Gene Sorter, shown below.



*Note: as of 8/31/2008, Genome-Wide Association Study data have been removed because the data producers (NIH and Wellcome Trust Case Control Consortium) have removed their data from public view due to privacy concerns.

Gene Sorter

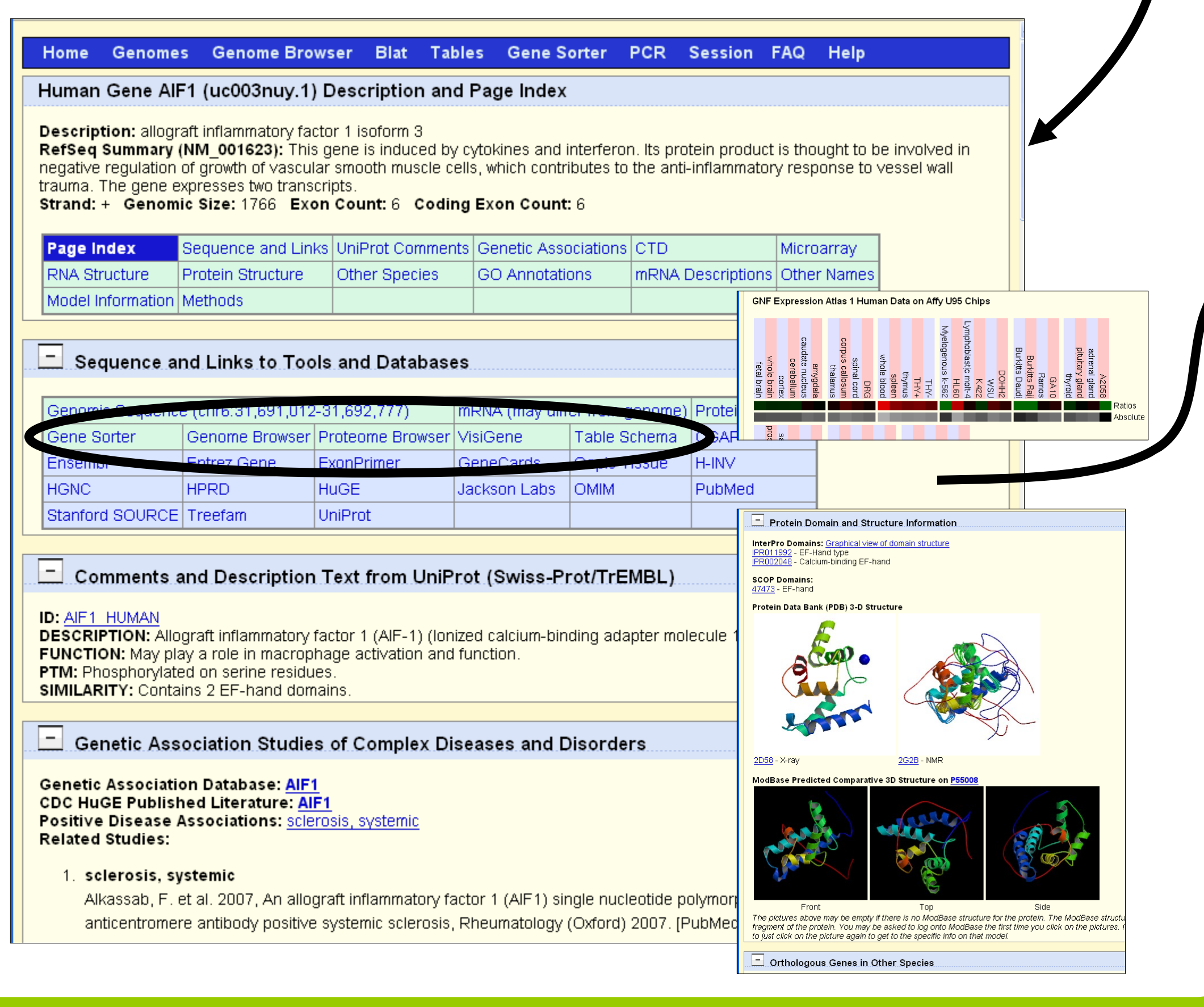
This highly configurable tool displays many types of data associated with genes, and genes can be reordered (or searched for) by expression, protein sequence similarity, similarity of Gene Ontology terms or PFAM domains, protein-protein interaction data, chromosomal distance or alphabetical sorting relative to the first gene in the list. Only a few of the many column types are shown here; the configure button leads to a page where the set of columns and their order can be changed. Filters can be defined to further narrow down the set of genes to those with the desired characteristics. The Description column contains links to Gene Details, shown below.



The list of genes to the right was obtained by selecting a score threshold of 10 in Genome Graphs and clicking the "sort genes" button, selecting AIF1 to be the reference gene, and selecting Expression (GNF Atlas 2) as the distance metric for sorting.

Gene Details

Information from many sources and links to other tools, both at UCSC and other institutions, are collected and displayed on the Gene Details page. The page to the right appears when clicking on the description of AIF in the Gene Sorter above, or on an AIF1 transcript image in the Genome Browser. To reduce the need for scrolling, the page begins with a table of links to its sections:



- Sequence and Links to Tools and Databases
- Comments and Description Text from UniProt (Swiss-Prot/TrEMBL)
- Genetic Association Studies of Complex Diseases and Disorders
- Comparative Toxicogenomics Database (CTD)
- Microarray Expression Data
- mRNA Secondary Structure of 3' and 5' UTRs
- Protein Domain and Structure Information
- Orthologous Genes in Other Species
- Gene Ontology (GO) Annotations with Structured Vocabulary
- Descriptions from all associated GenBank mRNAs
- Other Names for This Gene
- Gene Model Information
- Methods, Credits, and Use Restrictions

References

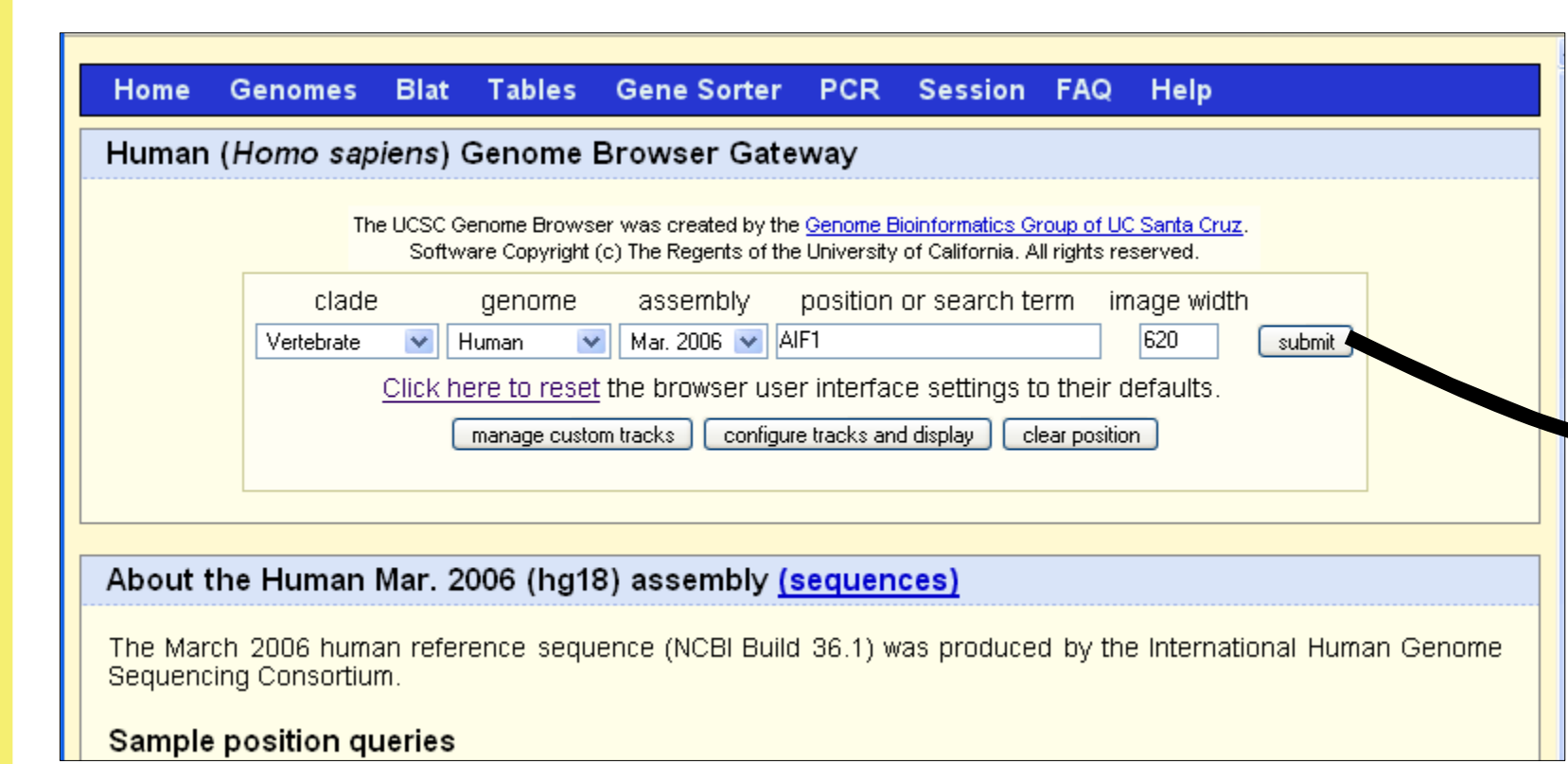
The UCSC Genome Browser Database: 2008 update. Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, Kober KM, Miller W, Pedersen JS, Pohl A, Raney BJ, Rhead B, Rosenbloom KR, Smith KE, Stanke M, Thakkapallayil A, Trumbower H, Wang T, Zweig AS, Haussler D, Kent WJ. Center for Biomolecular Science and Engineering, University of California Santa Cruz (UCSC). Santa Cruz, CA, 95064, USA.

Acknowledgements

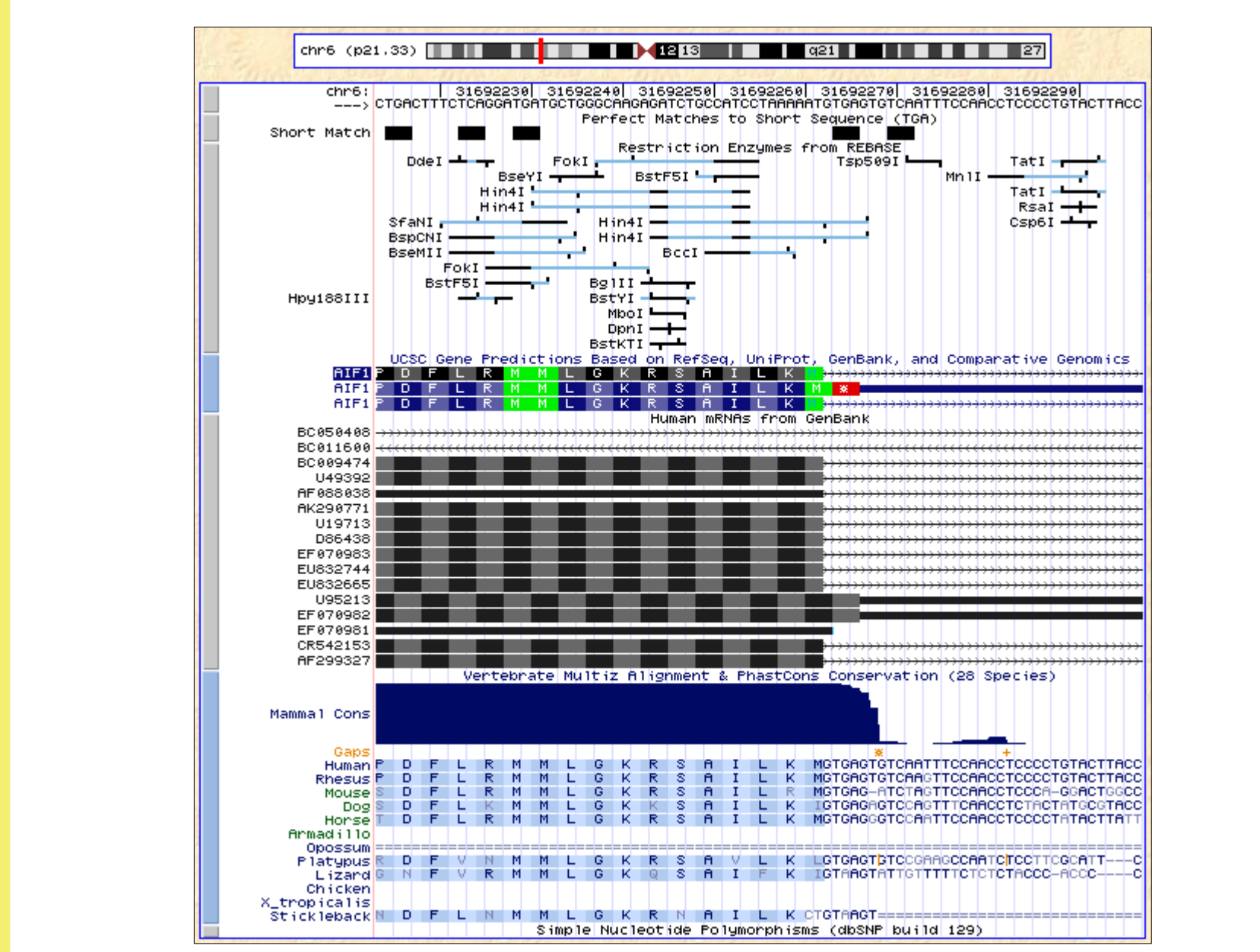
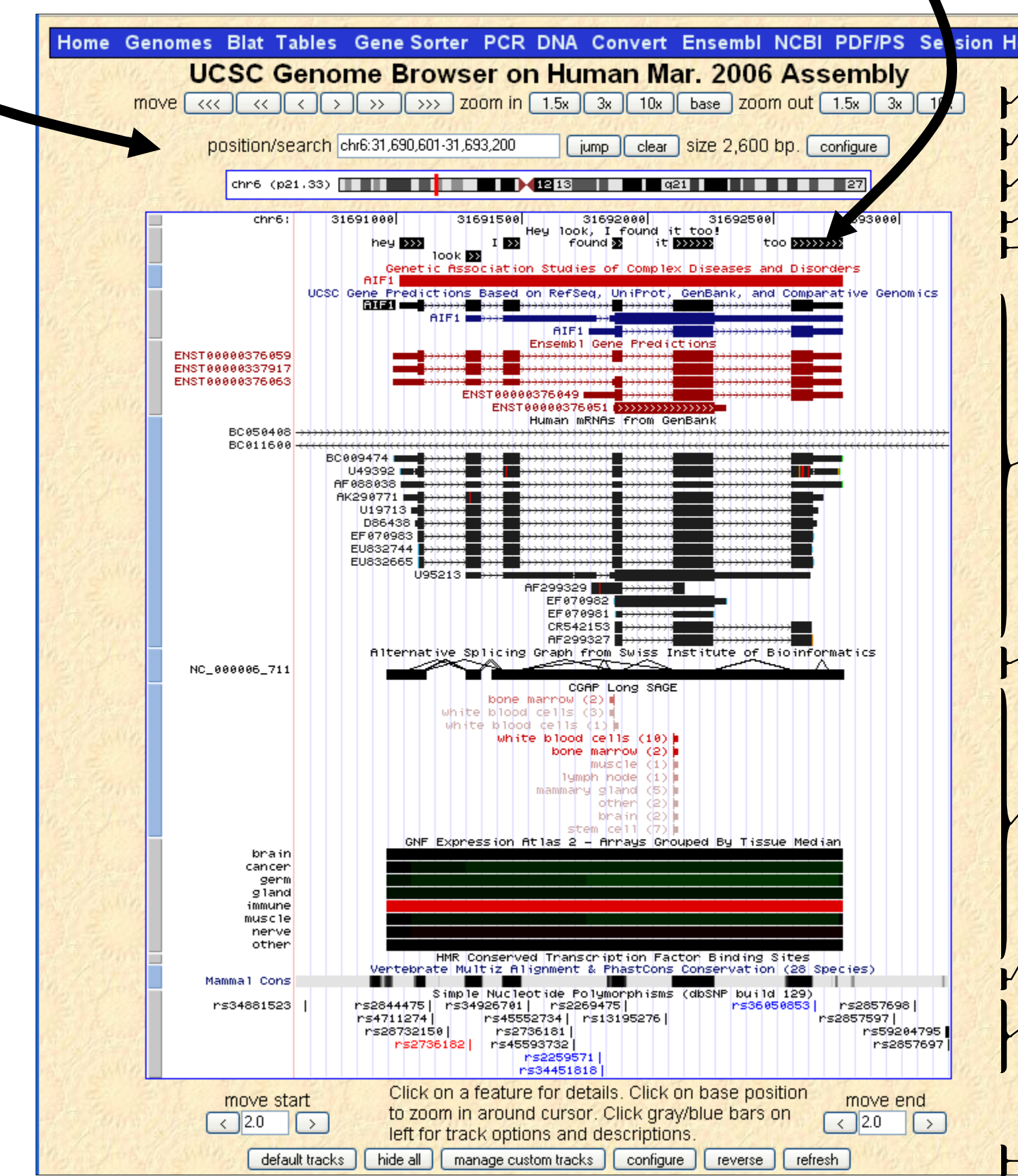
This work was funded by National Human Genome Research Institute (2 P41 HG002371-06 to UCSC Center for Genomic Science, 3 P41 HG002371-06 to UCSC ENCODE supplement to UCSC Center for Genomic Science); National Cancer Institute (Contract No. N01-CO-12400 for Mammalian Gene Collection).

Genome Browser

The Genome Browser provides a highly configurable and interactive window into a large collection of public genomic data. A region of the genome, of any size from a single base pair up to a chromosome, is selected by entering genomic coordinates, an accession or keyword(s) into the position/search text input. A plot of that region appears with genomic coordinates displayed along the top of the image. Below that, each visible dataset is displayed as a horizontal "track" of data. There are over 200 tracks available for the most recent human genome assembly (NCBI build 36, March 2006), so most are hidden by default. Tracks that are not hidden have multiple visibility modes: dense mode compresses all items into a single row, full mode gives each item its own row, pack mode puts multiple items in the same row when they fit, and squish mode is like pack but omits item labels and halves the row height. Clicking on a track in dense mode expands it into pack mode; clicking on an item in pack, squish or full mode leads to a page with a detailed description of the item, links to other sources of information when available, and a description of the track data, display conventions, methods and credits. Some tracks offer additional controls, e.g. filtering, display parameters, subtrack selection for tracks composed of similar datasets from the same source. Each track's control page can be reached in several ways: by clicking on the track's name in the table below the image or on the configuration page (configure button), by clicking on the light blue or gray button on the far left of the track image, or by clicking the track controls link on an item details page.



Easily upload your own data for viewing!



Enriched graphical features at base-level view: matches to short sequences, restriction enzymes, codons/codon differences.

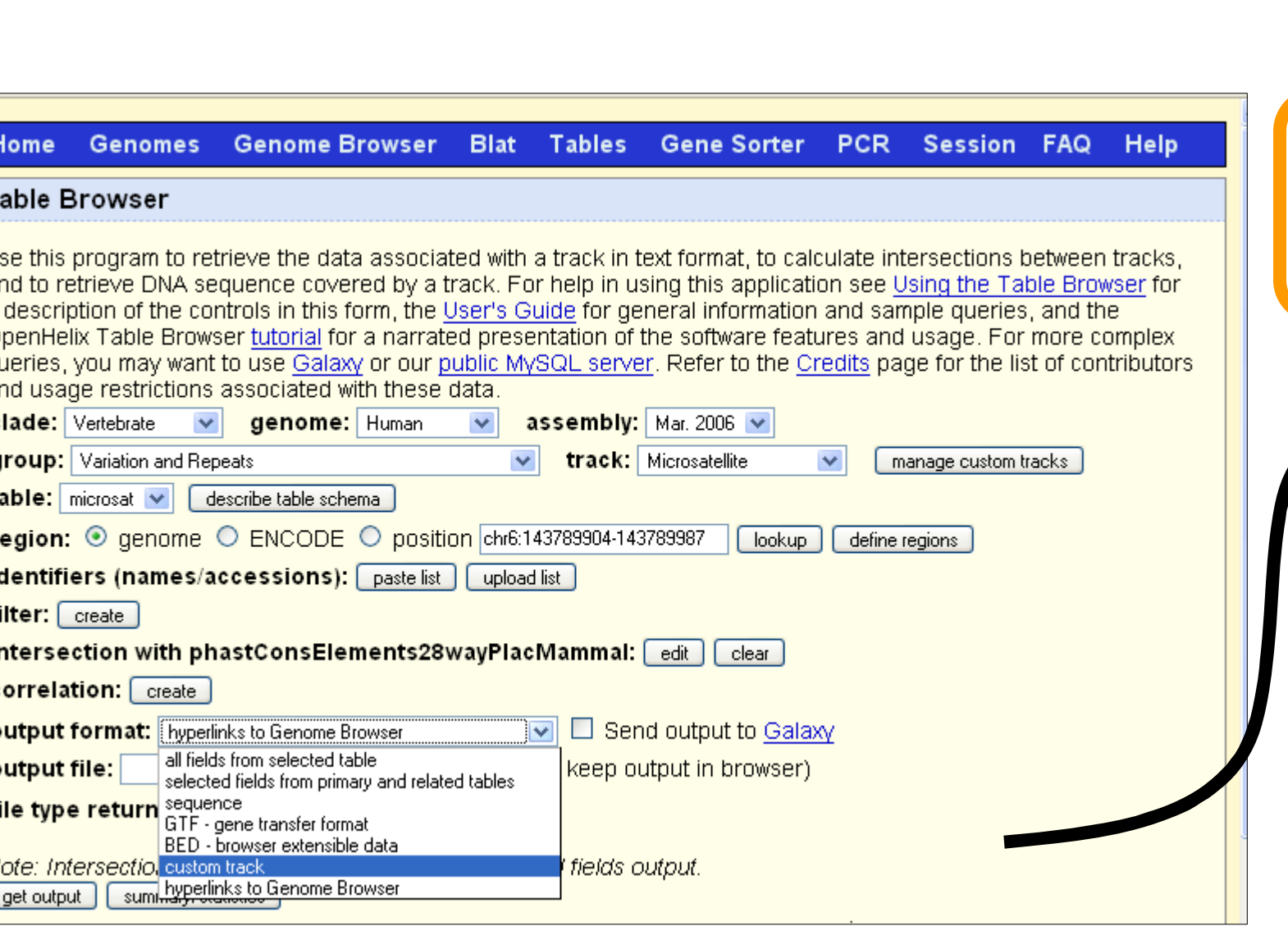
Gene-level view with uploaded data, disease association data, gene annotations, mRNAs, alt-splice, expression, conservation level, SNPs colored by function

- Links to other tools (UCSC and beyond)
- Adjust size and position of viewing window
- Enter position, accessions, keywords
- Configure appearance and select tracks
- Viewing window (red) on chromosome
- Base position: click to zoom in & re-center
- Uploaded data ("custom track"); >>> arrows indicate forward strand
- Gene predictions and cDNAs aligned to genome: tall blocks are coding exons, thin blocks are UTR, connecting lines are introns, arrows indicate forward or reverse strand; click to see more information
- Condensed alternative splicing display; click to see detailed graph
- Expression data: SAGE and microarray
- Condensed display of conservation score: darker regions are higher-scoring
- dbSNP reference SNPs: black is intergenic or intron, red is coding-nonsynonymous, blue is UTR; click to see more information including alignment
- Track controls; reverse button changes viewed strand forward to reverse; refresh accepts changes to individual track controls below image

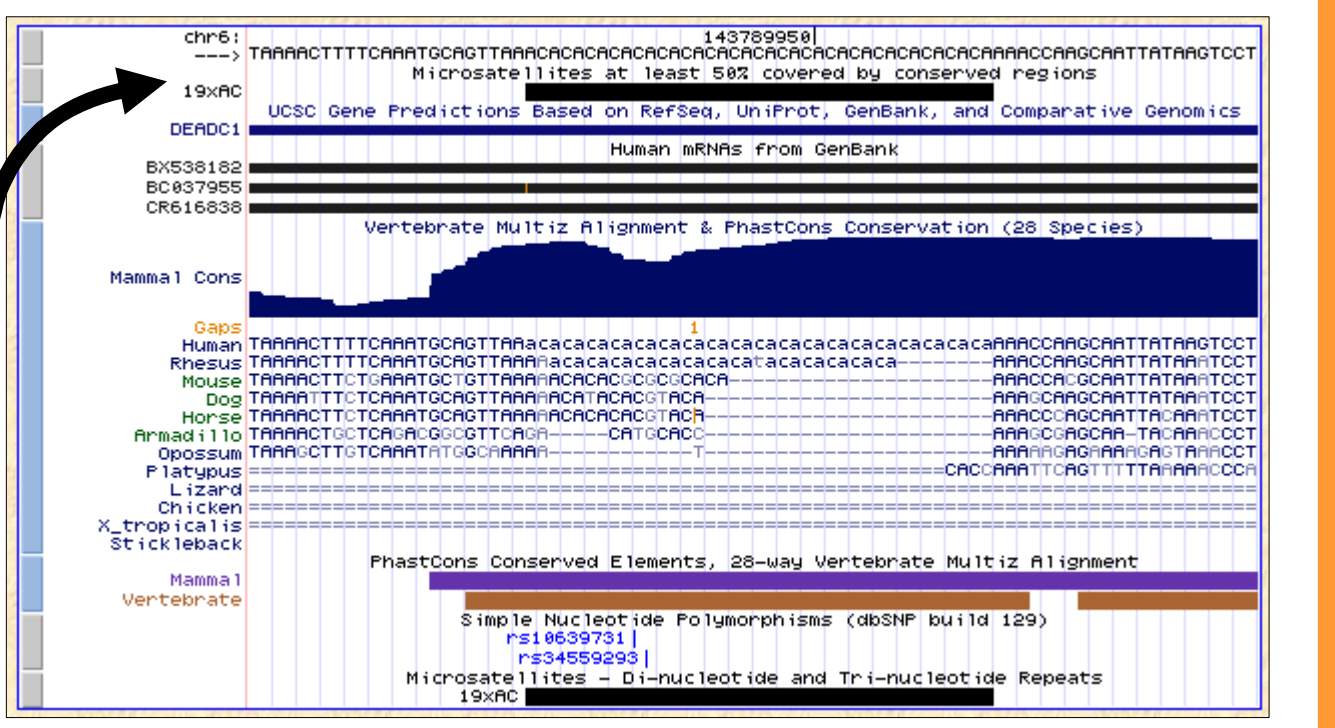
Table Browser

The Table Browser provides direct access to the contents of the database tables underlying the Genome Browser display. It supports simple queries such as downloading the entire contents of a table as text, as well as more complex queries that include filtering based on data values and intersection of two tracks by position. In addition to text results, query results can be sent to the Galaxy tool (<http://galaxy.psu.edu>) for even more complex queries, or loaded directly into the current session as custom tracks that will appear in the Genome Browser display and can be used in subsequent Table Browser queries.

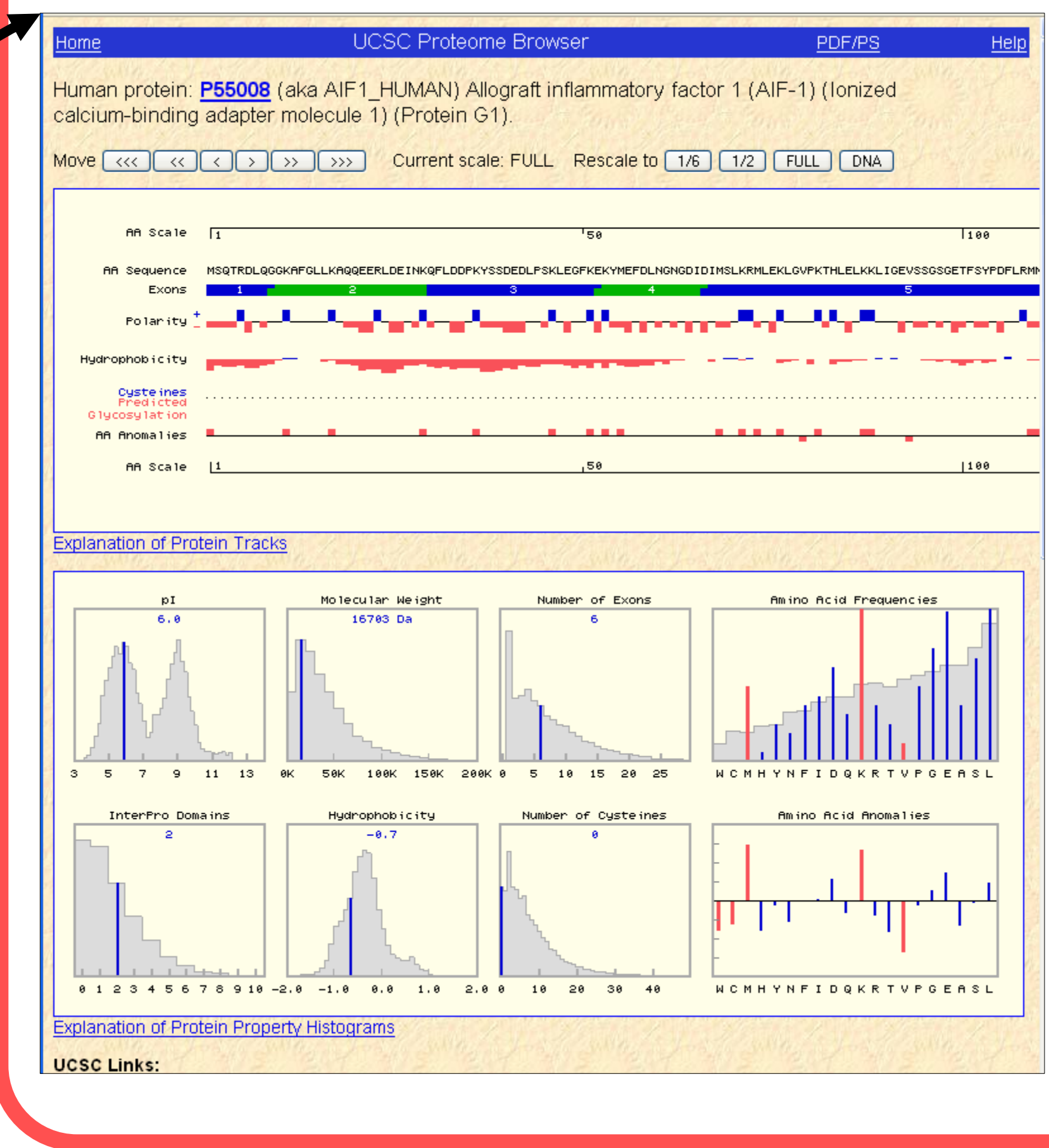
group, track: select dataset
manage custom tracks: add, delete, update uploaded data
 table: some tracks have auxiliary tables in addition to the main table that associates item data with genomic position
describe table schema: column descriptions and example data values
region: genome = data on genomes, ENCODE pilot regions (human only), or genomic position (can enter accessions/keywords here)
identifiers: upload list of identifiers that identify items in track's main table
filter: include only items that match specified criteria
intersection: include/exclude items by positional overlap with another track
correlation: calculate Pearson's r on data values of overlapping items of two tracks



Save query results as custom tracks that can be viewed in the Genome Browser and used for further Table Browser queries!



Proteome Browser



The Proteome Browser displays sequence-associated features of proteins obtained from UniProt. Like the Genome Browser, it plots annotations along the protein's amino acid sequence: exon boundaries, polarity, hydrophobicity, cysteines, predicted glycosylation sites, and amino acid anomalies (over- or underrepresented amino acids compared to frequencies across all proteins). Below the plot, several properties of the protein as a whole are superimposed on histograms summarizing the properties of all proteins in the database: isoelectric point, molecular weight, exon count, amino acid frequencies, InterPro domains, hydrophobicity, number of cysteines, and amino acid anomalies. Below the graphs are links to other UCSC tools, InterPro and SCOP, images from PDB and ModBase, and the protein sequence.

Other UCSC Tools

- Blat** swiftly aligns submitted nucleotide or protein sequences to the reference genome assembly.
- isPCR**, based on Blat, finds amplicons given primer sequences.
- VisiGene** provides a Google Maps-like interface to high-resolution mouse embryo in situ images.
- Sessions** enables saving and sharing of track views and configuration settings.
- Utilities:** batch coordinate conversion, sequence formatting cleaners, phylogenetic tree drawing tool.
- Downloads:** flatfiles for entire database contents.

More Information

Click "Help" link (upper right) to get tool-specific help pages. Click the "Contact Us" link (lower left of home page) to search for answers to questions sent to genome@soe.ucsc.edu, or email your question to the actively monitored public list. **OpenHelix** provides free training material: http://www.openhelix.com/downloads/ucsc/ucsc_home.shtml and also offers training seminars (some free or discounted). <http://genomewiki.ucsc.edu/> is a user-editable website that contains additional writings by UCSC staff and power-users.