UCSC Genome Bioinformatics

Kate Rosenbloom
Center for Biomolecular Science and Engineering
University of California, Santa Cruz

GMOD User Interface Caucus

January 18, 2007
About the UCSC Genome Bioinformatics Site

This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of in situ mouse and frog images to examine expression patterns.

News

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list.

1 January 2007 - Upcoming Genome Browser Seminars: SF, Seattle, NYC, Cleveland

The UCSC Bioinformatics Group announces four regional seminars and hands-on computer workshops on the UCSC Genome Browser, presented by OpenHelix:

- San Francisco, CA -- Wednesday, 31 January
- Seattle, WA -- Thursday, 1 February
- New York City, NY -- Tuesday, 13 February
- Cleveland, OH -- Wednesday, 14 February

Two sessions will be offered for the New York and Cleveland seminars: 9:00 a.m. to noon and 1 p.m. to 4 p.m. Only the afternoon session will be offered in San Francisco and Seattle.

http://genome.ucsc.edu
The UCSC Genome Browser Presents
Fully Annotated Genomes

Vertebrates
- human
- chimp
- rhesus macaque
- dog
- cow
- mouse
- rat
- opossum
- chicken
- tetraodon, fugu, zebrafish

Invertebrates
- sea squirt
- sea urchin
- fruitfly (12)
- honeybee
- mosquito
- worm (2)
- yeast

And coming soon...
- cat
- platypus
- medaka, stickleback
Hardware

• Under the hood

KiloKluster = 1000 CPUs
-- Linux Red Hat 9, Apache, Parasol
-- 10-Gigabit data transmission
-- dual 866 MHz machines x 500
-- 1 Gb RAM each

Smaller Clusters
-- 100-node cluster: dual Xeon 2.6 GHz
-- 400-node cluster

NFS
-- 12 machines on RAID arrays
-- 4 - 8 Gb RAM
-- 20+ Tb storage

• Public Site

-- 8 machines -- redundant
-- 64-bit
-- 8 Gb RAM
-- 1500 Gb storage
+ 15 blat servers
Data Contributors

- Human Genome Project
- Genbank/DDJ/EMBL contributors
- ENCODE Consortium
- Novartis GNF foundation
- Affymetrix, Perlegen, SNP Consortium
- SwissProt, Ensembl, EBI and NCBI
- Jackson Labs, RGD, Wormbase, Flybase
- Many contributors of gene prediction and other tracks.
High volume data handling

- All Genbank mRNAs loaded and aligned to the genome nightly; all ESTs weekly (24-48 hours to process).

- At least 6000 - 7000 regular users (separate IP addresses daily).

- 2 - 3 million hits a week

- Consistently #1 or #2 user of bandwidth on the UCSC campus
UCSC Bioinformatics Tools

- Genome Browser
- Table Browser
- Gene Sorter
- VisiGene
- Custom Tracks
- BLAT
- Downloads server, DAS server, mySQL access
Track configuration & description

Vertebrate Multiz Alignment & Conservation (17 Species)

Display mode: [pack] [Submit]

Pairwise alignments:
- *human*
- *chimp*
- *rhesus*
- *mouse*
- *rat*
- *rabbit*
- *dog*
- *cow*
- *armadillo*
- *elephant*
- *tenrec*
- *opossum*

Vertebrate
- *chicken*
- *X. tropicalis*
- *zebrafish*
- *tetraodon*
- *fugu*

Multiple alignment base-level:
- Display bases identical to reference as dots
- Display gaps between alignments

Codon Translation:
Default species to establish reading frame: [hg18]
- No codon translation
- Use default species reading frames for translation
- Use reading frames for species if available, otherwise use default species

Description

This track shows a measure of evolutionary conservation in 17 vertebrates, including mammalian, amphibian, bird, and fish species, based on a phylogenetic hidden Markov model, phastCons (Siepel et al., 2005). Multiz alignments of the following assemblies were used to generate this track:

- human (Mar. 2006, hg18)
- chimp (Nov 2003, panTro1)
- macaque (Jan 2006, rheMac2)
- mouse (Feb 2006, mm8)

Codon translation uses the following gene tracks as the basis for translation, depending on the species chosen:

<table>
<thead>
<tr>
<th>Gene Track</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known Genes</td>
<td>human, mouse, rat</td>
</tr>
<tr>
<td>RefSeq Genes</td>
<td>chicken</td>
</tr>
<tr>
<td>MGC Genes</td>
<td>X. tropicalis</td>
</tr>
<tr>
<td>Ensembl Genes</td>
<td>Fugu, chimp</td>
</tr>
<tr>
<td>mRNA</td>
<td>rhesus, rabbit, dog, cow, zebrafish</td>
</tr>
<tr>
<td>not translated</td>
<td>armadillo, elephant, tenrec, opossum, Tetraodon</td>
</tr>
</tbody>
</table>

Methods

Best-in-genome pairwise alignments were generated for each species using blastz, followed by chaining and netting. The pairwise alignments were then multiply aligned using multiz, following the ordering of the species tree diagrammed above. The resulting multiple alignments were then assigned conservation scores by phastCons, using a tree model with branch lengths derived from the ENCODE project Multi-Species Sequence Analysis group, September 2005 tree model. This tree was generated from TBA alignments over 23 vertebrate species and is based on 4D sites.

The phastCons program computes conservation scores based on a phylo-HMM, a type of probabilistic model that describes both the process of DNA substitution at each site in a genome and the way this process changes from one site to the next (Felsenstein and Churchill 1996, Yang 1995, Siepel and...
Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See Using the Table Browser for a description of the controls in this form. For more complex queries, you may want to use our public MySQL server. Refer to the Credits page for the list of contributors and usage restrictions associated with these data.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>group:</td>
<td>Genes and Gene Prediction Tracks</td>
<td>track:</td>
<td>Known Genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>table:</td>
<td>knownGene</td>
<td>describe table schema</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>region:</td>
<td></td>
<td>position: chr7:113842512-114117390</td>
<td>lookup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>identifiers (names/accessions):</td>
<td>paste list</td>
<td>upload list</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>filter:</td>
<td>create</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intersection:</td>
<td>create</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correlation:</td>
<td>create</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>output format:</td>
<td>selected fields from primary and related tables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>output file:</td>
<td>(leave blank to keep output in browser)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>file type returned:</td>
<td>plain text</td>
<td>gzip compressed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To reset all user cart settings (including custom tracks), click here.
<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>VisiGene</th>
<th>whole brain</th>
<th>small intestine</th>
<th>large intestine</th>
<th>kidney</th>
<th>ovary</th>
<th>pancreas</th>
<th>heart</th>
<th>lung</th>
<th>liver</th>
<th>BLASTP E-Value</th>
<th>Rankprop Score</th>
<th>Genome Position</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BRCA1</td>
<td>I446</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>0</td>
<td>1</td>
<td>chr17 38,490,250</td>
</tr>
<tr>
<td>2</td>
<td>ITGB3BP</td>
<td>97924</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>0.00234442</td>
<td>chr1 63,659,669</td>
<td>replication factor C3 isoform 1</td>
</tr>
<tr>
<td>3</td>
<td>RFC3</td>
<td>161999</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>0.210275</td>
<td>chr13 33,299,938</td>
<td>replication factor C3 isoform 1</td>
</tr>
<tr>
<td>4</td>
<td>Y15758</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr13 51,938,044</td>
<td>L1 protein</td>
</tr>
<tr>
<td>5</td>
<td>RACGAP1</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr12 48,687,350</td>
<td>Rac GTPase activating protein 1</td>
</tr>
<tr>
<td>6</td>
<td>NUP107</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr12 67,394,868</td>
<td>nucleoporin 107KDa</td>
</tr>
<tr>
<td>7</td>
<td>MRPL39</td>
<td>99488</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr21 25,890,756</td>
<td>mitochondrial ribosomal protein L39</td>
</tr>
<tr>
<td>8</td>
<td>POLE2</td>
<td>100419</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr14 49,202,356</td>
<td>DNA polymerase epsilon subunit 2</td>
</tr>
<tr>
<td>9</td>
<td>RFC4</td>
<td>100881</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>0.21743</td>
<td>chr3 187,998,689</td>
<td>replication factor C4</td>
</tr>
<tr>
<td>10</td>
<td>A290612</td>
<td>102240</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr15 39,436,457</td>
<td>Nucleolar and spindle associated protein</td>
</tr>
<tr>
<td>11</td>
<td>DSCR2</td>
<td>83207</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr21 39,473,284</td>
<td>Down syndrome critical region protein</td>
</tr>
<tr>
<td>12</td>
<td>ZWINT</td>
<td>54581</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr10 57,789,122</td>
<td>ZW10 interactor isoform a</td>
</tr>
<tr>
<td>13</td>
<td>RBBP8</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>0.402232</td>
<td>chr18 18,814,521</td>
<td>retinoblastoma binding protein 8</td>
</tr>
<tr>
<td>14</td>
<td>CR749951</td>
<td>101368</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr1 47,459,826</td>
<td>TAL1 (SCL) interrupting locus (Fragment)</td>
</tr>
<tr>
<td>15</td>
<td>GMNN</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr6 24,888,700</td>
<td>geminin</td>
</tr>
<tr>
<td>16</td>
<td>TRPP13</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>0.211022</td>
<td>chr5 958,581</td>
<td>thyroid hormone receptor interactor 13</td>
</tr>
<tr>
<td>17</td>
<td>KIF2C</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr1 44,888,575</td>
<td>kinesin family member 2C</td>
</tr>
<tr>
<td>18</td>
<td>KIF14</td>
<td>n/a</td>
<td>whole brain</td>
<td>bone marrow</td>
<td>thymus</td>
<td>bone-CD4+ T-cells</td>
<td>skin</td>
<td>spleen</td>
<td>pancreatic islets</td>
<td>heart</td>
<td>lung</td>
<td>liver</td>
<td>n/a</td>
<td>chr1 197,287,241</td>
<td>kinesin family member 14</td>
</tr>
</tbody>
</table>
Visigene (a “virtual microscope”)
About the ENCODE Project

This site contains information related to the ENCODE project at NHGRI. The UCSC Genome Bioinformatics Group manages the official repository of sequence-related data for the ENCODE consortium and supports the coordination of data submission, storage, retrieval, and visualization. A summary of the status of datasets submitted to UCSC by ENCODE contributors is available on the ENCODE data status page.

UCSC also has a special interest in comparative genomics, and we provide resources for the ENCODE multiple sequence alignment interest group. Ensembl also provides an ENCODE resource page.

We'd like to thank NHGRI and the contributors of annotations and analyses to this project. The team at UCSC that develops and maintains this ENCODE site is made up of Daryl Thomas, Kate Rosenbloom, Jim Kent, and the UCSC Genome Bioinformatics staff. Read more.

News

7 Oct. 2006 - Comparative Genomics Data Release

Twelve tracks of data produced by the ENCODE Multi-Species Sequence Analysis group have been released to the UCSC public server. These tracks contain multiple sequence alignments, conservation, and conserved (constrained) elements produced by four conservation methods (phastCons, binCons, GERP, SCONEx) applied to three sequence alignments (TBA, MLAGAN, MAVID), and also an assessment of the agreement among the alignment methods. The alignments were based on genomic sequence in the ENCODE regions of 28 vertebrate species, as defined in the MSA September 2005 sequence freeze.

http://genome.ucsc.edu/ENCODE
<table>
<thead>
<tr>
<th>Region</th>
<th>Description</th>
<th>Chr</th>
<th>Size (~Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENm001</td>
<td>CFTR</td>
<td>7</td>
<td>1.9</td>
</tr>
<tr>
<td>ENm002</td>
<td>Interleukin</td>
<td>5</td>
<td>1.0</td>
</tr>
<tr>
<td>ENm003</td>
<td>Apo Cluster</td>
<td>11</td>
<td>0.5</td>
</tr>
<tr>
<td>ENm004</td>
<td>Chr22 Pick</td>
<td>22</td>
<td>1.7</td>
</tr>
<tr>
<td>ENm005</td>
<td>Chr21 Pick</td>
<td>21</td>
<td>1.7</td>
</tr>
<tr>
<td>ENm006</td>
<td>ChrX Pick</td>
<td>X</td>
<td>1.2</td>
</tr>
<tr>
<td>ENm007</td>
<td>Chr19 Pick</td>
<td>19</td>
<td>1.0</td>
</tr>
<tr>
<td>ENm008</td>
<td>Alpha Globin</td>
<td>16</td>
<td>0.5</td>
</tr>
<tr>
<td>ENm009</td>
<td>Beta Globin</td>
<td>11</td>
<td>1.0</td>
</tr>
<tr>
<td>ENm010</td>
<td>HOXA Cluster</td>
<td>7</td>
<td>0.5</td>
</tr>
<tr>
<td>ENm011</td>
<td>IGFB2/H19</td>
<td>11</td>
<td>0.6</td>
</tr>
<tr>
<td>ENm012</td>
<td>FOXP2</td>
<td>7</td>
<td>1.0</td>
</tr>
<tr>
<td>ENm013</td>
<td>Manual</td>
<td>7</td>
<td>1.1</td>
</tr>
<tr>
<td>ENm014</td>
<td>Manual</td>
<td>7</td>
<td>1.2</td>
</tr>
<tr>
<td>ENr11</td>
<td>Random</td>
<td>13</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr12</td>
<td>Random</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr13</td>
<td>Random</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr14</td>
<td>Random</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr15</td>
<td>Random</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr16</td>
<td>Random</td>
<td>18</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr17</td>
<td>Random</td>
<td>12</td>
<td>0.5</td>
</tr>
<tr>
<td>ENr18</td>
<td>Random</td>
<td>2</td>
<td>0.5</td>
</tr>
</tbody>
</table>
New features: Genomewiki

http://genomewiki.cse.ucsc.edu
New features: Custom track manager

**Add Custom Tracks**

- **clade**: Vertebrate
- **genome**: Human
- **assembly**: Mar. 2006
- **hg18**

Display your own data as custom annotation tracks in the browser. Data must be formatted in BED, GFF, GTF, WIG, or PSL formats. To configure the display, set track and browser line attributes as described in the User's Guide. Publicly available custom tracks are listed here. Examples are here.

Optional track documentation:

Paste URLs or data: Or upload: Browse... Submit

Manage Custom Tracks

- **genome**: Human
- **assembly**: Mar. 2006
- **hg18**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Type</th>
<th>Doc</th>
<th>Items</th>
<th>Pos</th>
<th>delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>regulatory</td>
<td>TeleGene(tm) Regulatory Regions</td>
<td>gff</td>
<td>2</td>
<td>chr22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Click here for a

- add custom tracks
- go to genome browser
- go to table browser
New feature: Track reordering

Configure Tracks

Control tracks in all groups here: hide all | show all | default | Control track visibility more selectively below.

Mapping and Sequencing Tracks

<table>
<thead>
<tr>
<th>Track Order:</th>
<th>hide all</th>
<th>show all</th>
<th>default</th>
<th>submit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Position</td>
<td>dense</td>
<td>Chromosome position in bases. (Clicks here zoom in 3x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gap</td>
<td>hide</td>
<td>Gap Locations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAC End Pairs</td>
<td>hide</td>
<td>BAC End Pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC Percent</td>
<td>hide</td>
<td>Percentage GC in 20,000-Base Windows</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short Match</td>
<td>hide</td>
<td>Perfect Matches to Short Sequence (AAAAA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restr Enzymes</td>
<td>hide</td>
<td>Restriction Enzymes from REBASE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genes and Gene Prediction Tracks

<table>
<thead>
<tr>
<th>Track Order:</th>
<th>hide all</th>
<th>show all</th>
<th>default</th>
<th>submit</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlyBase Genes</td>
<td>pack</td>
<td>FlyBase Protein-Coding Genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RefSeq Genes</td>
<td>dense</td>
<td>RefSeq Genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB Noncoding</td>
<td>pack</td>
<td>FlyBase Noncoding Genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-SCAN</td>
<td>hide</td>
<td>N-SCAN Gene Predictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geneid Genes</td>
<td>hide</td>
<td>Geneid Gene Predictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genscan Genes</td>
<td>hide</td>
<td>Genscan Gene Predictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Augustus Genes</td>
<td>hide</td>
<td>Augustus Gene Predictions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Proteins</td>
<td>pack</td>
<td>Human(hg17) proteins mapped by chained tBLASTn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
New features: Comparative genomics

- Gap annotation
- Genomic breaks
- Codon translation at base level
New features (under review): Saving user sessions

Sign in to UCSC Genome Bioinformatics

Signing in enables you to save current settings into a named session, and then restore settings from the session later. If you wish, you can share named sessions with other users.

The sign-in page is handled by our wiki system [click here to sign in]. The wiki also serves as a forum for users to share knowledge and ideas.

Session Management

[Click here to reset] the browser user interface settings to their defaults.

If you [sign in], you will also have the option to save named sessions.

Load Settings

Load settings from another user's saved session:
user: [ ] [submit] session name: [ ]

Load settings from a local file: [ ] [Browse...] [submit]

Load settings from a URL (http://..., ftp://...): [ ] [submit]

Save Settings

Save current settings to a local file:
file: [ ] [submit]

(file type returned: [ ] [plain text] [submit])

(leave file blank to get output in browser window)
New features (in development): Whole genome graphing

SNP association study, prepublication data
GMOD Scenario #1: Search for gene by name…

About the Gene Sorter

This program displays a sorted table of genes that are related to one another. The relationship can be one of several types, including protein-level homology, similarity of gene expression profiles, or genomic proximity.
GMOD Scenario #1:  

... and view information page
GMOD Scenario #1:

... and view information page (2)
GMOD Scenario #1: 
... and view information page (3)

### Homologous Genes in Other Species (BLASTP Best Hit)

<table>
<thead>
<tr>
<th>Gene Details</th>
<th>Gene Details</th>
<th>Gene Details</th>
<th>Gene Details</th>
<th>Gene Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse Genome Browser</td>
<td>Mouse Genome Browser</td>
<td>Mouse Genome Browser</td>
<td>Mouse Genome Browser</td>
<td>Mouse Genome Browser</td>
</tr>
<tr>
<td>Rat Genome Browser</td>
<td>Rat Genome Browser</td>
<td>Rat Genome Browser</td>
<td>Rat Genome Browser</td>
<td>Rat Genome Browser</td>
</tr>
<tr>
<td>Zebrafish Gene Details</td>
<td>Zebrafish Gene Details</td>
<td>Zebrafish Gene Details</td>
<td>Zebrafish Gene Details</td>
<td>Zebrafish Gene Details</td>
</tr>
<tr>
<td>D. melanogaster Genome Browser</td>
<td>D. melanogaster Genome Browser</td>
<td>D. melanogaster Genome Browser</td>
<td>D. melanogaster Genome Browser</td>
<td>D. melanogaster Genome Browser</td>
</tr>
<tr>
<td>C. elegans Genome Browser</td>
<td>C. elegans Genome Browser</td>
<td>C. elegans Genome Browser</td>
<td>C. elegans Genome Browser</td>
<td>C. elegans Genome Browser</td>
</tr>
<tr>
<td>S. cerevisiae Genome Browser</td>
<td>S. cerevisiae Genome Browser</td>
<td>S. cerevisiae Genome Browser</td>
<td>S. cerevisiae Genome Browser</td>
<td>S. cerevisiae Genome Browser</td>
</tr>
</tbody>
</table>

### Jackson Lab Annotations

**RGD** (Ensembl FlyBase WormBase SGD)

### Gene Ontology (GO) Annotations with Structured Vocabulary

**GO:00034676** nucleic acid binding

**GO:00034677** DNA binding

**GO:0003701** transcription factor activity

**GO:0003565** sequence-specific DNA binding

**GO:0046872** metal ion binding

**Biological Process:**

**GO:0006550** transcription

**GO:0006555** regulation of transcription, DNA-dependent

**Cellular Component:**

**GO:0005634** nucleus

### Descriptions from all associated GenBank mRNAs

- **AK131266** - Homo sapiens cDNA FLJ16201 fs, clone CTONG2008721, highly similar to Homo sapiens CAGHH4 mRNA.
- **AJ454870** - Homo sapiens forkhead transcription factor (POXP2) mRNA, partial cds; alternatively spliced.
- **AF467252** - Homo sapiens clone FC5 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **BC118104** - Homo sapiens forkhead box P2, mRNA (cDNA clone IMAGE:4285527), complete cds.
- **AY144815** - Homo sapiens brain forkhead/winged helix transcription factor FOXP2 isoform mRNA, complete cds; alternatively spliced.
- **AF467253** - Homo sapiens clone FOXP2 short isoform (FOXP2) mRNA, complete cds.
- **AF524840** - Homo sapiens putative forkhead/winged-helix transcription factor (FOXP2) mRNA, complete cds.
- **AY467257** - Homo sapiens clone HFC2 forkhead/winged helix transcription factor (FOXP2) mRNA, complete cds; alternatively spliced.
- **AJ454870** - Homo sapiens forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **USO741** - Homo sapiens CAGHH4 mRNA, partial cds.
- **AF467253** - Homo sapiens clone AMYG 2a.2 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **AF467254** - Homo sapiens clone AMYG 4a.4 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **BC104134** - Homo sapiens forkhead box P2, mRNA (cDNA clone MGC:161382 IMAGE:899320), complete cds.
- **AF467253** - Homo sapiens clone 3RACE700 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **AF467256** - Homo sapiens clone BA4 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **AF467254** - Homo sapiens clone STR7 forkhead/winged helix transcription factor (FOXP2) mRNA, partial cds; alternatively spliced.
- **AJ454870** - Homo sapiens full length insert cDNA clone YY325E7.
- **DQ778626** - Homo sapiens forkhead box P2 variant 3 mRNA, complete cds, alternatively spliced.
GMOD Scenario #2 (sort of):
Search by keyword

<table>
<thead>
<tr>
<th>clade</th>
<th>genome</th>
<th>assembly</th>
<th>position or search term</th>
<th>image width</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertebrate</td>
<td>Human</td>
<td>Mar. 2006</td>
<td>zinc ion binding</td>
<td>620</td>
</tr>
</tbody>
</table>

Click here to reset the browser user interface settings to their defaults.

manage custom tracks | configure tracks and display | clear position

**Known Genes**

- PAPPAs (AF142993) at chr1:7460639-17507859
- PAPPAs (NH_020248) at chr1:7460590-17507559
- PAPPAs (NH_020247) at chr1:7460590-17507559
- PAPPAs (NH_020240) at chr1:7460639-17507859
- PAPPAs (NH_020245) at chr1:7460632-17507859
- PAPPAs (NH_020249) at chr1:7460639-17507859
- PAPPAs (NH_020246) at chr1:7460639-17507859
- PAPPAs (NH_020241) at chr1:7460639-17507859
- PAPPAs (NH_020242) at chr1:7460639-17507859
- PAPPAs (NH_020243) at chr1:7460639-17507859
- PAPPAs (NH_020244) at chr1:7460639-17507859

**About the Human Mar. 2006 (hg18) assembly (sequences)**

The March 2006 human reference sequence (NCBI Build 36.1) was produced by the International Human Genome Sequencing Consortium.

**Sample position queries**

A genome position can be specified by the accession number of a sequenced genomic clone, an EST or STS marker, or a cytological band, a chromosomal coordinate range, or keywords from the description of an mRNA. The following list shows examples of valid position queries for the human genome build hg18. See the User's Guide for more information.

**Non-Human RefSeq Genes**

- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047
- ATG2D4 (AF124026) at chr1:41660382-44605047

GMOD Scenario #2 (sort of):
Search by keyword

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz. Software Copyright (c) The Regents of the University of California. All rights reserved.
GMOD Scenario #3: Customized report on aspects of gene
GMOD Scenario #3 Alternate:
Customized report on aspects of gene

- Exon count
- GO terms
- Swiss-Prot disease description
GMOD Scenario #3: Customized report on gene, cont.
GMOD Scenario #3: Report on aspects of gene, cont.(2)

- Exon count
- GO terms
- Swiss-Prot disease description

hg18.kgXref fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>kgID</td>
<td>Known Gene ID</td>
</tr>
<tr>
<td>mRNA</td>
<td>mRNA ID</td>
</tr>
<tr>
<td>spID</td>
<td>SWISS-PROT protein Accession number</td>
</tr>
<tr>
<td>spDisplayID</td>
<td>SWISS-PROT display ID</td>
</tr>
<tr>
<td>geneSymbol</td>
<td>Gene Symbol</td>
</tr>
<tr>
<td>refseq</td>
<td>RefSeq ID</td>
</tr>
<tr>
<td>protAcc</td>
<td>NCBI protein Accession number</td>
</tr>
<tr>
<td>description</td>
<td>Description</td>
</tr>
</tbody>
</table>

proteome.spDisease fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>accession</td>
<td>SWISS-PROT accession number</td>
</tr>
<tr>
<td>displayID</td>
<td>SWISS-PROT display ID</td>
</tr>
<tr>
<td>diseaseDesc</td>
<td>disease description</td>
</tr>
</tbody>
</table>

Select Fields from hg18.knownGene

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Name of gene</td>
</tr>
<tr>
<td>chrom</td>
<td>Reference sequence chromosome or scaffold</td>
</tr>
<tr>
<td>strand</td>
<td>+ or - for strand</td>
</tr>
<tr>
<td>txStart</td>
<td>Transcription start position</td>
</tr>
<tr>
<td>txEnd</td>
<td>Transcription end position</td>
</tr>
<tr>
<td>cgStart</td>
<td>Coding region start</td>
</tr>
<tr>
<td>cgEnd</td>
<td>Coding region end</td>
</tr>
<tr>
<td>exonCount</td>
<td>Number of exons</td>
</tr>
<tr>
<td>exonStarts</td>
<td>Exon start positions</td>
</tr>
<tr>
<td>exonEnds</td>
<td>Exon end positions</td>
</tr>
<tr>
<td>proteinID</td>
<td>SWISS-PROT ID</td>
</tr>
<tr>
<td>alignID</td>
<td>Unique identifier for each (known gene, alignment position) pair</td>
</tr>
</tbody>
</table>

go.term fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td></td>
</tr>
<tr>
<td>name</td>
<td></td>
</tr>
<tr>
<td>term_type</td>
<td></td>
</tr>
<tr>
<td>acc</td>
<td></td>
</tr>
<tr>
<td>is_obsolete</td>
<td></td>
</tr>
<tr>
<td>is_root</td>
<td></td>
</tr>
</tbody>
</table>

Phs18.knownGene.name  | hg18.knownGene.exonCount  | go.term.name | hg18.kgXref.geneSymbol | proteome.spDisease.diseaseDesc
CR749236             | 23                      | n/a          | FOXO2                   | n/a
CR749236             | 23                      | nucleo acid binding, nucleus, zinc ion binding, FOXO2 | n/a
NC010181             | 3                       | n/a          | FOXO2                   | n/a
AV119615              | 18                      | nucleo acid binding, DNA binding, transcription factor activity, nucleus, transcription, regulation of transcription, DNA dependent, zinc ion binding | n/a
AV119615              | 18                      | n/a          | FOXO2                   | n/a
NM_116991             | 18                      | nucleo acid binding, DNA binding, transcription factor activity, nucleus, transcription, regulation of transcription, DNA dependent, zinc ion binding | n/a
NM_116991             | 17                      | nucleo acid binding, DNA binding, transcription factor activity, nucleus, transcription, regulation of transcription, DNA dependent, zinc ion binding | n/a
GMOD Scenarios 4 & 5: Bulk queries and external data integration; Compare user gene set to UCSC Known Genes

- How many user genes are not in Known Genes?
- How well conserved across different species are the genes unique to the user gene set?
GMOD Scenarios 4 & 5: Loading external data

Add Custom Tracks

clade: vertebrate, genome: Human, assembly: Mar. 2006, hg18

Display your own data as custom annotation tracks in the browser. Data must be formatted in BED, GFF, GTF, WIG or PSL formats. To configure the display, set track and browser line attributes as described in the User's Guide. Publicly available custom tracks are listed here. Examples are here.

Paste URL or data:

http://hgdev.cse.ucsc.edu/~kate/test/my GENES. gtf

Optional track documentation:

This is a gene track (actually Ensembl genes for hg18).

Click here for an HTML document template that may be used for Genome Browser track descriptions.

Manage Custom Tracks

genome: Human, assembly: Mar. 2006, hg18

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Type</th>
<th>Doc</th>
<th>Items</th>
<th>Pos</th>
<th>delete</th>
<th>update</th>
<th>add custom tracks</th>
<th>go to genome browser</th>
<th>go to table browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>My GENES</td>
<td>My Gene Track (really Ensembl)</td>
<td>gff</td>
<td>Y</td>
<td>59069</td>
<td>chr3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GMOD Scenarios 4 & 5: Loading external data, cont.
GMOD Scenarios 4 & 5: Intersection on whole dataset

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See Using the Table Browser for a description of the controls in this form. For more complex queries, you may want to use our public MySQL server. Refer to the Credits page for the list of contributors and usage restrictions associated with these data.

- clade: vertebrate
- genome: Human
- assembly: Mar. 2006
- group: Custom Tracks
- track: My Genes
- table: ct, MyGenes
- remove custom track
- describe table schema
- region: genome position chr3:185536414-185546755
- identifiers (names/accessions)
- filter: create
- intersection: create
- correlation: create
- output format: selected fields from primary and related tables
- output file:
- file type returned: plain text

Intersect with My Genes

Select a group, track and table to intersect with:

- group: Genes and Gene Prediction Tracks
- track: Known Genes
- table: Known Genes (knownGene)

These combinations will maintain the gene/alignment structure (if any) of My Genes:

- All My Genes records that have any overlap with Known Genes
- All My Genes records that have no overlap with Known Genes
- All My Genes records that have at least 80% overlap with Known Genes
- All My Genes records that have at most 80% overlap with Known Genes
GMOD Scenarios 4 & 5: Intersection on whole dataset, cont.

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. See Using the Table Browser for a description of the controls in this form. For more complex queries, you may want to use our public MySQL server. Refer to the Credits page for the list of contributors and usage restrictions associated with these data.

- clade: Vertebrate
- group: Custom Tracks
- table: ct_MyGenes
- region: genome 1 1 1
- identifiers (names/accessions): paste list
- correlation: create
- intersection with knownGene: edit
- output format: custom track
- output file: (leave blank)
- file type returned: plain text

Note: Intersection doesn't work with all fields or selected fields.

Output ct_MyGenes as Custom Track

- Custom track header:
  - name: My Unknown Genes
  - description: My Genes Not In Known Genes
  - visibility: pack
- url:

Create one BED record per:
- Whole Gene
- Upstream by 200 bases
- Exons plus 0 bases at each end
- Introns plus 0 bases at each end
- 5' UTR Exons
- Coding Exons
- 3' UTR Exons
- Downstream by 200 bases

Note: If a feature is close to the beginning or end of a chromosome and upstream/downstream added, they may be truncated in order to avoid extending past the edge of the chromosome.
Kent’s UI Guidelines

- Keep it reliable
- Keep it fast
- Label everything in plain English
- Put the most commonly used controls on the top of the page
- Keep it as simple as possible (but no simpler)
- Try to make options work together in an orthogonal way
- Remember your users are *intelligent* professionals. Don’t dumb things down; complexity comes with the territory
- Don’t change the site unnecessarily once people have gotten used to it.
User interface challenges: User-configurable ordering

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>Item Number in Displayed List/Select Gene</td>
<td>n/a</td>
</tr>
<tr>
<td>Name</td>
<td>Gene Name/Select Gene</td>
<td>n/a</td>
</tr>
<tr>
<td>UniProt</td>
<td>UniProt (SwissProt/TrEMBL) Protein Display ID</td>
<td>n/a</td>
</tr>
<tr>
<td>UniProt Acc</td>
<td>UniProt (SwissProt/TrEMBL) Protein Accession</td>
<td>n/a</td>
</tr>
<tr>
<td>RefSeq</td>
<td>NCBI RefSeq Gene Accession</td>
<td>n/a</td>
</tr>
<tr>
<td>Entrez Gene</td>
<td>NCBI Entrez Gene/Link ID</td>
<td>n/a</td>
</tr>
<tr>
<td>GenBank</td>
<td>GenBank mRNA Accession</td>
<td>n/a</td>
</tr>
<tr>
<td>Ensembl</td>
<td>Ensembl Transcript ID</td>
<td>n/a</td>
</tr>
<tr>
<td>GNF Atlas 2 ID</td>
<td>ID of Associated GNF Atlas 2 Expression Data</td>
<td>n/a</td>
</tr>
<tr>
<td>VisoGene</td>
<td>UCSC VisoGene In Situ Image Browser</td>
<td>n/a</td>
</tr>
<tr>
<td>Allen Brain</td>
<td>Allen Brain Atlas In Situ Images of Adult Mouse Brains</td>
<td>n/a</td>
</tr>
<tr>
<td>U133 ID</td>
<td>ID of Associated Affymetrix U133 Expression Data</td>
<td>n/a</td>
</tr>
<tr>
<td>U133Plus2 ID</td>
<td>ID of Associated Affymetrix U133 Plus 2.0 Expression Data</td>
<td>n/a</td>
</tr>
<tr>
<td>U95 ID</td>
<td>ID of Associated Affymetrix U95 Expression Data</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Configure Tracks

Control tracks in all groups here: hide all | show all | default | submit Control track visibility more selectively below.

<table>
<thead>
<tr>
<th>Mapping and Sequencing Tracks Track Order:</th>
<th>hide all</th>
<th>show all</th>
<th>default</th>
<th>submit Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Position dense</td>
<td>Chromosome position in bases. (Clicks here zoom in 3x)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gap</td>
<td>Gap Locations</td>
<td>11</td>
<td>map</td>
<td></td>
</tr>
<tr>
<td>BAC End Pairs</td>
<td>BAC End Pairs</td>
<td>15</td>
<td>map</td>
<td></td>
</tr>
<tr>
<td>GC Percent</td>
<td>Percentage GC in 20,000-Base Windows</td>
<td>23</td>
<td>map</td>
<td></td>
</tr>
<tr>
<td>Short Match</td>
<td>Perfect Matches to Short Sequence (AAAAA)</td>
<td>99</td>
<td>map</td>
<td></td>
</tr>
<tr>
<td>Resr Enzymes</td>
<td>Restriction Enzymes from REBASE</td>
<td>99.9</td>
<td>map</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genes and Gene Prediction Tracks Track Order:</th>
<th>hide all</th>
<th>show all</th>
<th>default</th>
<th>submit Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlyBase Genes pack</td>
<td>FlyBase Protein-Coding Genes</td>
<td>34</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>RefSeq Genes dense</td>
<td>RefSeq Genes</td>
<td>35</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>FB Noncoding pack</td>
<td>FlyBase Noncoding Genes</td>
<td>35</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>N-SCAN hide</td>
<td>N-SCAN Gene Predictions</td>
<td>45.1</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>Geneid Genes hide</td>
<td>Geneid Gene Predictions</td>
<td>49</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>Genscan Genes hide</td>
<td>Genscan Gene Predictions</td>
<td>50</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>Augustus Genes hide</td>
<td>Augustus Gene Predictions</td>
<td>51.7</td>
<td>genes</td>
<td></td>
</tr>
<tr>
<td>Human Protds pack</td>
<td>Human(hg17) proteins mapped by chained tBLASTn</td>
<td>142</td>
<td>genes</td>
<td></td>
</tr>
</tbody>
</table>
User interface challenges:
Track grouping to avoid overload
User interface challenges: Composite tracks to group similar data

Affymetrix ChIP/Chip (retinoic acid-treated HL-60 cells) Sites

Display mode: hide submit

Select subtracks:
- All add clear
- Timepoint:
  - 0hrs add clear
  - 2hrs add clear
  - 8hrs add clear
  - 32hrs add clear
- Factor:
  - Brg1 add clear
  - CELF6 add clear
  - CTCF add clear
  - H3K27me3 add clear
  - H4Kac4 add clear
  - P300 add clear
  - PU1 add clear
  - RARA add clear
  - Pol2 add clear
  - SIRT1 add clear
  - TFIIB add clear

Show checkboxes for: Only selected subtracks All subtracKs
- Affymetrix ChIP/Chip (Brg1 retinoic acid-treated HL-60, 0hrs) Sites
- Affymetrix ChIP/Chip (Brg1 retinoic acid-treated HL-60, 2hrs) Sites
- Affymetrix ChIP/Chip (Brg1 retinoic acid-treated HL-60, 8hrs) Sites
User Support and Training

• FAQs:  http://genome.cse.ucsc.edu/FAQ/
• questions?  genome@soe.ucsc.edu
  archived answers:
    http://genome.ucsc.edu/contacts.html
• OpenHelix:  http://www.openhelix.com/
  – Classes, seminars
  – Free online tutorial
  – Quick reference cards
Thanks!

• UCSC Genome Browser Team:
  – David Haussler – PI
  – Jim Kent – Browser Concept, BLAT, Team Leader
  – Donna Karolchik – Engineering Mgr, Docs & Training
  – Mark Diekhans, Fan Hsu, Angie Hinrichs, Kate Rosenbloom, Hiram Clawson, Rachel Harte, Heather Trumbower, Galt Barber, Andy Pohl - Engineering
  – Robert Kuhn (mgr), Ann Zweig, Kayla Smith, Brooke Rhead, Archana Thakkapallayil – QA/Support

• Funding:
  – National Human Genome Research Institute
  – Howard Hughes Medical Institute
  – National Cancer Institute