Exploring relationships and mining data with the UCSC Gene Sorter Supplementary Data

Table R.1 All columns currently available in the Gene Sorter for the human gene set.

Column Name	Column Description
Name	Gene Name/Select Gene
Swiss-Prot	Swiss-Prot Protein Display ID
SP Acc	Swiss-Prot Protein Accession
RefSeq	NCBI RefSeq Gene Accession
LocusLink	NCBI LocusLink ID
GenBank	GenBank mRNA Accession
Ensembl	Ensembl Transcript ID
HInv	HInv ID
GNF Atlas 2 ID	ID of Associated GNF Atlas 2 Expression Data
U133Plus2 ID	ID of Associated Affymetrix U133 Plus 2.0 Expression Data
U133 ID	ID of Associated Affymetrix U133 Expression Data
U95 ID	ID of Associated Affymetrix U95 Expression Data
GNF Atlas 2	GNF Expression Atlas 2 Data from U133A and GNF1H Chips
Max GNF Atlas 2	Maximum Expression Value of GNF Expression Atlas 2
GNF Atlas 2 Delta	Normalized Difference in GNF Expression Atlas 2 from Selected Gene
GNF U95	GNF Expression Atlas 1 Human Data on Affy U95 Chips
Max GNF U95	Maximum Expression Value of GNF Expression Atlas 1
GNF U95 Delta	Normalized Difference in GNF Atlas 1 Expression from Selected Gene
GNF Atlas1 Delta	Normalized Difference in GNF Atlas 1 Expression from Selected Gene
BLASTP Bits	NCBI Blastp Bit Score
BLASTP E-Value	NCBI Blastp E-Value
%ID	NCBI Blastp Percent Identity
5' UTR Fold	5' UTR Fold Energy (Estimated kcal/mol)
3' UTR Fold	3' UTR Fold Energy (Estimated kcal/mol)
Genome Position	Genome Position/Link to Genome Browser
Coding SNPs	Simple Nucleotide Polymorphisms in Coding Regions
Exon Count	Number of Exons (Including Non-Coding)
Mouse	Mouse Ortholog (Best Blastp Hit to UCSC Known mm3)
Rat	Rat Ortholog (Best Blastp Hit to UCSC Known rn3)
Zebrafish	*Danio rerio* Ortholog (Best Blastp Hit to Ensembl)
Drosophila	*D. melanogaster* Ortholog (Best Blastp Hit to FlyBase Proteins)
C. elegans	*C. elegans* Ortholog (Best Blastp Hit to WormPep)
Yeast	*Saccharomyces cerevisiae* Ortholog (Best Blastp Hit to RefSeq)
Pfam Domains	Protein Family Domain Structure
Superfamily	Protein Superfamily Assignments
PDB	Protein Data Bank
Gene Ontology	Gene Ontology (GO) Terms Associated with Gene
Description	Short description Line/Link to Details Page

Table R.2 All sort options currently available in the Gene Sorter for the human gene set.

Sort Type	Sort Description
Expression (GNF Atlas2)	Difference in expression with selected gene according to GNF Gene Expression Atlas2
Expression (GNF Atlas1)	Difference in expression with selected gene according to GNF Gene Expression Atlas1
Protein Homology - BLASTP	BLASTP E-value with selected gene
Protein Homology - Rankprop	Rankprop similarity score with selected gene
Protein Homology - PSI-BLAST	BLASTP E-value with selected gene
Pfam Similarity	Number of Pfam domains shared with selected gene
Gene Distance	Distance in base pairs from selected gene
Chromosome	Ordered by chromosome position
Name Similarity	Number of leading characters in name that match selected gene name
Alphabetical	Alphabetical order of name
GO Similarity	Number of shared gene ontology terms

Table M.1 Column types in columnDb.ra. The first six column types are general purpose; there are many columns of each of these types. The remaining column types, which are more specialized, have only a single instance of each.

Type	Function
lookup	Displays a string for each gene based on a simple database table lookup. Only a single value is associated with each gene.
association	Displays multiple strings for each gene based on an arbitrary SQL query
distance	Looks up and displays a floating point number based on two genes – the selected gene and the gene for the current row.
expRatio	Shows microarray expression ratios.
expMulti	Shows microarray data as absolute or expression.
num	Displays the row number in the table. Usually the leftmost column.
knownName	Displays biological name of gene if available. Clicking here selects gene.
acc	Displays the gene ID.
knownPos	Displays position of gene in genome. Hyperlinks to Genome Browser.
knownDetails	Displays a sentence about gene. Hyperlinks to details page.
pfam	Shows Pfam domain. Can filter on domain name or Pfam ID.
go	Shows GO term. Can filter on term or term accession.

Table M.2 Sort types in orderDb.ra. The expression similarity and protein homology based sorts are instances of the pair type. Pfam and GO similarity are instances of the association type.

Type	Function
pair	Looks up floating point distance value in database table using ID of selected gene and ID of gene to sort.
group	Partitions genes into non-overlapping groups. Genes in same group as selected gene have a distance of one; other genes have an infinite distance.
association	Distance is based on the number of terms in an association table that the gene shares with the selected gene.
geneDistance	Distance is the number of bases between middle of gene and middle of selected gene. Genes on other chromosomes have infinite distance.
nameSimilarity	Distance is based on the number of letters -- starting from the beginning of the biological gene name – that are the same between the gene and the selected gene.
abc	Sorts genes alphabetically by biological name. The selected gene is irrelevant in this ordering.
genomePos	Sorts genes by chromosome and position on chromosome. The selected gene is irrelevant in the ordering.

Table M.3 The methods of the column object.

Method	Purpose
exists	Returns true if all tables necessary to display column are available.
cellVal	Returns string value for a particular gene.
cellPrint	Prints out cell in html table for a particular gene.
labelPrint	Prints column label in html table.
tableColumns	Returns how many html columns this logical column uses. (Expression data sets with many tissues use multiple html columns for readability.)
configControls	Prints out controls for this column on configuration page.
simpleSearch	Called when user inputs text in search box. Returns a searchResult structure for each gene that matches search in this column.
filterControls	Prints out controls for this column on filter page.
advFilter	Return a subset of the genes that passes the filter in this column.