

## Schema for UCSC Genes - UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

**Database:** mm9 **Primary Table:** knownGene **Row Count:** 55,419

**Format description:** Genes based on RefSeq, GenBank, and UniProt.

| field      | example          | SQL type         | info                   | description   |
|------------|------------------|------------------|------------------------|---|
| name       | uc007aet.1       | varchar(255)     | <a href="#">values</a> | Name of gene  |
| chrom      | chr1             | varchar(255)     | <a href="#">values</a> | Reference sequence chromosome or scaffold   |
| strand     | -                | char(1)          | <a href="#">values</a> | + or - for strand   |
| txStart    | 3195984          | int(10) unsigned | <a href="#">range</a>  | Transcription start position  |
| txEnd      | 3205713          | int(10) unsigned | <a href="#">range</a>  | Transcription end position  |
| cdsStart   | 3195984          | int(10) unsigned | <a href="#">range</a>  | Coding region start   |
| cdsEnd     | 3195984          | int(10) unsigned | <a href="#">range</a>  | Coding region end   |
| exonCount  | 2                | int(10) unsigned | <a href="#">range</a>  | Number of exons   |
| exonStarts | 3195984,3203519, | longblob         |                        | Exon start positions  |
| exonEnds   | 3197398,3205713, | longblob         |                        | Exon end positions  |
| proteinID  |                  | varchar(40)      | <a href="#">values</a> | UniProt display ID for Known Genes, UniProt accession or RefSeq protein ID for UCSC Genes |
| alignID    | uc007aet.1       | varchar(255)     | <a href="#">values</a> | Unique identifier for each (known gene, alignment position) pair                          |