UCSC Genes

Genecats 3/20/13

Brian Raney

Talk Overview

- What is UCSC genes?
- Where do the gene models come from?
- How does it all plug into the Genome Browser?

- Interactive Q&A

UCSC Genes Overview

- Semi-conservative set of gene models
- Stable id's with version numbers
- Built on latest mouse and human every six months(?)
- ~25 supporting tables (kg*, known*)
- Gene models based on several lines of evidence
  - RefSeq, mRNA, EST, Rfam, CCDS, MAFs, exoniphy
- Extensive link out resources
- Questionable Future
- Two supporting databases
  - proteome, uniProt

Method Sketch

- Build uniProt database
- Build proteome database
- Build gene models
- Build link-out tables

uniProt Database

- comprehensive, high-quality and freely accessible resource of protein sequence and functional information.
- ftp://ftp.expasy.org/databases/uniprot/current_release/knowledgebase/complete/
  - uniprot_sprot.dat.gz
  - uniprot_trembl.dat.gz
  - uniprot_sprot_varsplic.fasta.gz
- 37 tables
- Doc in kent/src/hg/makeDb/doc/uniProt

Proteome Database

- Local mish-mash of remote sourced data and cross-reference tables.
- HGNC (HUGO Gene Nomenclature Committee)
- Reactome (biological pathways)
- Interpro (protein familes)
- Pfam (protein domains)
- Doc in kent/src/hg/makeDb/doc/proteins
- Probably should be retired.

Build Gene Models

- Grab Input data
  - Genbank alignments
  - CCDS (genbank)
  - tRNA (Lowe lab)
  - Rfam (rna families, Sanger)
  - Exoniphy (Siepel, deprecated)

Build Gene Models (2)

- Separate out antibody regions
- Build splicing graphs for mRNA and ESTs
- Establish evidence weights
  - refSeq 100
  - ccds 50
  - mrna 2
  - txOrtho 1
  - exoniphy 1
  - est 1
- Create EST evidence when 2 native ESTs support it

Build Gene Models(3)

- Map xDb mRNAs to our species for txOrtho
- Do the txWalk!

- Tools in kent/src/hg/txGraph
- Doc in kent/src/hg/makeDb
  - /doc/ucscGenes
- More detail on trackUi page
  - http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=knownGene

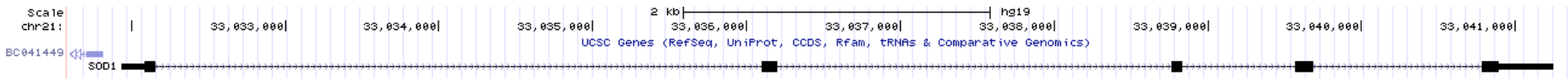
Methods

The UCSC Genes are built using a multi-step pipeline:

1. RefSeq and GenBank RNAs are aligned to the genome with BLAT, keeping only the best alignments for each RNA. Alignments are discarded if they do not meet certain sequence identity and coverage filters. All sequences must align with high (98%) identity. The sequence coverage must be at least 90% for shorter sequences (those with 2500 or fewer bases), with the coverage threshold progressively relaxed for longer sequences.
2. Alignments are broken up at non-intronic gaps, with small isolated fragments thrown out.
3. A splicing graph is created for each set of overlapping alignments. This graph has an edge for each exon or intron, and a vertex for each splice site, start, and end. Each RNA that contributes to an edge is kept as evidence for that edge. Gene models from the Consensus CDS project (CCDS) are also added to the graph.
4. A similar splicing graph is created in the mouse, based on mouse RNA and ESTs. If the mouse graph has an edge that is orthologous to an edge in the human graph, that is added to the evidence for the human edge.
5. If an edge in the splicing graph is supported by two or more human ESTs, it is added as evidence for the edge.
6. If there is an Exoniphy prediction for an exon, that is added as evidence.
7. The graph is traversed to generate all unique transcripts. The traversal is guided by the initial RNAs to avoid a combinatorial explosion in alternative splicing. All RefSeq transcripts are output. For other multi-exon transcripts to be output, an edge supported by at least one additional line of evidence beyond the RNA is required. Single-exon genes require either two RNAs or two additional lines of evidence beyond the single RNA.
8. Alignments are merged in from the hg19 tRNA Genes track and from Rfam in regions that are syntenic with the mm9 mouse genome.
9. Protein predictions are generated. For non-RefSeq transcripts we use the txCdsPredict program to determine if the transcript is protein-coding, and if so, the locations of the start and stop codons. The program weighs as positive evidence the length of the protein, the presence of a Kozak consensus sequence at the start codon, and the length of the orthologous predicted protein in other species. As negative evidence it considers nonsense-mediated decay and start codons in any frame upstream of the predicted start codon. For RefSeq transcripts the RefSeq protein prediction is used directly instead of this procedure. For CCDS proteins the CCDS protein is used directly.
10. The corresponding UniProt protein is found, if any.
11. The transcript is assigned a permanent "uc" accession. If the transcript was not in the previous release of UCSC Genes, the accession ends with the suffix ".1" indicating that this is the first version of this transcript. If the transcript is identical to some transcript in the previous release of UCSC Genes, the accession is re-used with the same version number. If the transcript is not identical to any transcript in the previous release but it overlaps a similar transcript with a compatible structure, the previous accession is re-used with the version number incremented.

Tables

- knownGene
  - The central genepred format table.
- kgColor
  - RGB colors for knownGene display in hgTracks



Gene Clusters

- knownCanonical
  - Clustered knownGene to (mostly) have one gene per locus
- knownIsoforms
  - Mapping of canonical cluster id's to knownGene id's
- Used in gene suggest (?)

UCSC Gene Sequence Tables

- knownGeneMrna
  - Sequence of mRNA associated with every gene model.
- knownGenePep
  - Protein sequence from the CDS of translated sequence in knownGeneMrna
- knownGeneTxMrna
  - The mRNA sequence if it had been transcribed from the reference genome
- knownGeneTxPep
  - Protein sequence from from the CDS of translated sequence in knownGeneMTxrna

knownAlt

- Alt splicings, output from txgAnalyze, (91,222)
- altThreePrime, bleedingExon, retainedIntron, cassetteExon, altFivePrime, altPromoter, altFinish, strangeSplice, atacIntron,
- Bed6
  - chr1 12594 12612 altThreePrime 0 +
  - chr1 1440 16606 bleedingExon 0 -
- Used?

Blast Tables

- Built by doHgNearBlastp.pl
- Sum of all in knownBlastTab (updated?)
- Other species
  - Mouse knownGene mmBlastTab
  - Rat rgdGene rnBlastTab
  - Worm sangerGene ceBlastTab
  - Zebrafish Ensembl drBlastTab
  - Human knownGene hgBlastTab
  - Fly flyBaseGene dmBlastTab

BlastTab Schema

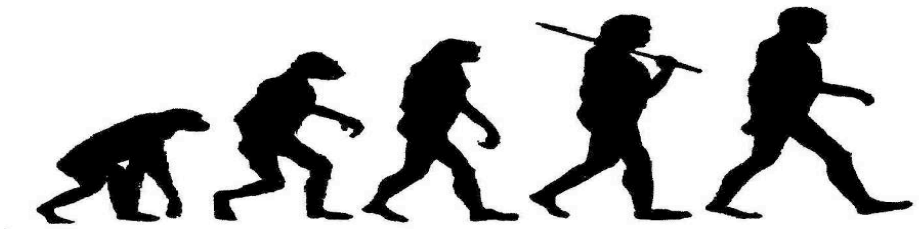
Field	Type	Null	Key	Default	Extra
query	varchar(255)	NO	MUL	NULL	
target	varchar(255)	NO		NULL	
identity	float	NO		NULL	
aliLength	int(10) unsigned	NO		NULL	
mismatch	int(10) unsigned	NO		NULL	
gapOpen	int(10) unsigned	NO		NULL	
qStart	int(10) unsigned	NO		NULL	
qEnd	int(10) unsigned	NO		NULL	
tStart	int(10) unsigned	NO		NULL	
tEnd	int(10) unsigned	NO		NULL	
eValue	double	NO		NULL	
bitScore	double	NO		NULL	

g

Mappings from previous set

- If previous set on different assembly built with pslMap
- Names like kg5ToKg6
- Status from: exact, compatible, overlap, none

oldId	oldChrom	oldStart	oldEnd	newId	status	note
uc001aaa.3	chr1	11873	14409	uc001aaa.3	exact	
uc010nxq.1	chr1	11873	14409	uc010nxq.1	exact	
uc010nxr.1	chr1	11873	14409	uc010nxr.1	exact	



kgXref

Field	Type	Null	Key	Default	Extra
kgID	varchar(255)	NO	MUL	NULL	
mRNA	varchar(255)	NO	MUL	NULL	
spID	varchar(255)	NO	MUL	NULL	
spDisplayID	varchar(255)	NO	MUL	NULL	
geneSymbol	varchar(255)	NO	MUL	NULL	
refseq	varchar(255)	NO	MUL	NULL	
protAcc	varchar(255)	NO	MUL	NULL	
description	longblob	NO		NULL	
rfamAcc	varchar(255)	NO	MUL	NULL	
tRnaName	varchar(255)	NO	MUL	NULL	

Misc Tables

- kgAlias
  - Aliases for knownGene from uniProt
- kgProtAlias
  - Mapping of knownGene to uniProt displayId and name
- kgSpAlias
  - Mapping of knownGene ids to uniProt
- kgProtMap2
  - CDS exons in PSL format
- kgTargetAli
  - PSL mapping of knownGeneTxMrna sequence to reference genome

Overlap Tables

- knownToAllenBrain
  - Mapping of Allen Brain Atlas id's to knownGene id's
- knownToDecipher
  - Built by DECIPHER build process, overlaps of knownGene with DECIPHER table
- knownCanonToDecipher
  - Built by DECIPHER load process, maps overlap of knownCanonical table and Decipher table
- knownToEnsembl
  - Overlap of Ensembl and knownGene. Built by knownGene and Ensembl? process

Overlap Tables (2)

- knownToGnfAtlas2
- knownToHInv
- knownToHprd
- knownToKeggEntrez
- knownToLocusLink
- knownToPfam
- knownToRefSeq
- knownToSuper
- knownToTreefam
- knownToU133
- knownToVisiGene

Human Gene SOD1 (uc002ypa.3) Description and Page Index

Description: Homo sapiens superoxide dismutase 1, soluble (SOD1), mRNA.
RefSeq Summary (NM_000454): The protein encoded by this gene binds copper and zinc ions and is one of two isozymes responsible for destroying free superoxide radica
molecular oxygen and hydrogen peroxide. The other isozyme is a mitochondrial protein. Mutations in this gene have been implicated as causes of familial amyotrophic later
the publications that are available for this gene. Please see the Gene record to access additional publications. ##RefSeq-Attributes-START## Transcript_exon_combination
Transcript (Including UTRs):
Position: chr21:33,031,935-33,041,243 **Size:** 9,309 **Total Exon Count:** 5 **Strand:** +
Coding Region:
Position: chr21:33,032,083-33,040,891 **Size:** 8,809 **Coding Exon Count:** 5

Page Index	Sequence and Links	UniProtKB Comments	Genetic Associations	CTD	Microarray
RNA Structure	Protein Structure	Other Species	GO Annotations	mRNA Descriptions	Pathways
Other Names	GeneReviews	Model Information	Methods		

Data last updated: 2011-12-21

Sequence and Links to Tools and Databases

Genomic Sequence (chr21:33,031,935-33,041,243)			mRNA (may differ from genome)		Protein (154 aa)
Gene Sorter	Genome Browser	Protein FASTA	VisiGene	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	GeneNetwork
Gepis Tissue	H-INV	HGNC	HPRD	Human Cortex Gene Expression	Jackson Lab
MOPED	OMIM	PubMed	Reactome	Stanford SOURCE	Treefam
UniProtKB	Wikipedia				

Comments and Description Text from UniProtKB

ID: SODC_HUMAN
DESCRIPTION: RecName: Full=Superoxide dismutase [Cu-Zn]; EC=1.15.1.1; AltName: Full=Superoxide dismutase 1; Short=hSod1;
FUNCTION: Destroys radicals which are normally produced within the cells and which are toxic to biological systems.
CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
COFACTOR: Binds 1 copper ion per subunit.
COFACTOR: Binds 1 zinc ion per subunit.
SUBUNIT: Homodimer; non-disulfide linked. Homodimerization may take place via the ditryptophan cross-link at Trp-33. The pathogenic variants ALS1 Arg-38, Arg-47, Arg-8
protein does not.
INTERACTION: P26339:Chga (xeno); NbExp=5; IntAct=EBI-990792, EBI-990900; P16014:Chgb (xeno); NbExp=6; IntAct=EBI-990792, EBI-990820;
SUBCELLULAR LOCATION: Cytoplasm. Note=The pathogenic variants ALS1 Arg-86 and Ala-94 gradually aggregates and accumulates in mitochondria.
PTM: Unlike wild-type protein, the pathogenic variants ALS1 Arg- 38, Arg-47, Arg-86 and Ala-94 are polyubiquitinated by RNF19A leading to their proteasomal degradation. T
PTM: The ditryptophan cross-link at Trp-33 is responsible for the non-disulfide-linked homodimerization. Such modification might only occur in extreme conditions and additi
DISEASE: Defects in SOD1 are the cause of amyotrophic lateral sclerosis type 1 (ALS1) [MIM:105400]. ALS1 is a familial form of amyotrophic lateral sclerosis, a neurodeg
years. The etiology of amyotrophic lateral sclerosis is likely to be multifactorial, involving both genetic and environmental factors. The disease is inherited in 5-10% of case
MISCELLANEOUS: The protein (both wild-type and ALS1 variants) has a tendency to form fibrillar aggregates in the absence of the intramolecular disulfide bond or of boun
SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
WEB RESOURCE: Name=Alsod; Note=ALS genetic mutations db; URL="http://alsod.iop.kcl.ac.uk/Als/";
WEB RESOURCE: Name=GeneReviews; URL="http://www.ncbi.nlm.nih.gov/sites/GeneTests/lab/gene/SOD1";
WEB RESOURCE: Name=NIEHS-SNPs; URL="http://egp.gs.washington.edu/data/sod1/";
WEB RESOURCE: Name=Wikipedia; Note=Superoxide dismutase entry; URL="http://en.wikipedia.org/wiki/Superoxide_dismutase";

Genetic Association Studies of Complex Diseases and Disorders

Genetic Association Database: SOD1
CDC HuGE Published Literature: SOD1
Positive Disease Associations: ALS/amyotrophic lateral sclerosis , amyotrophic lateral sclerosis cause novel protein interactions , familial amyotrophic lateral sclerosis , f
Related Studies:

1. **ALS/amyotrophic lateral sclerosis**
   Mancuso, M. et al. 2002, A screening for superoxide dismutase-1 D90A mutation in Italian patients with sporadic amyotrophic lateral sclerosis., Amyotrophic lateral scl
   Our results confirm that recessive D90A mutation is present in Italy and it is associated with the phenotype already described A screening for that mutation, easily ma
2. **amyotrophic lateral sclerosis cause novel protein interactions**
   Kunst CB et al. 1997, Mutations in SOD1 associated with amyotrophic lateral sclerosis cause novel protein interactions., Nature genetics, 1997 Jan;15(1):91-4. [PubM

hgGeneData

- kent/src/hg/hgGene/hgGeneData (cgi-bin/hgGeneData)
  - C_elegans
  - D_melanogaster
  - Human
  - Mouse
  - Rat
  - S_cerevisiae
  - Zebrafish
- Assembly subdirs
- genome.ra
- links.ra
- otherOrgs.ra
- section.ra

Genome.ra

name global

knownGeneMrna knownGeneMrna

summaryTables refSeqSummary knownToRefSeq

summarySql select summary from refSeqSummary,knownToRefSeq where
knownToRefSeq.name='%s' and (refSeqSummary.mrnaAcc=knownToRefSeq.value)

summaryIdSql select value from knownToRefSeq where name='%s'

summarySource RefSeq Summary

Description (genome.ra)

Description: Homo sapiens superoxide dismutase 1, soluble (SOD1), mRNA.
RefSeq Summary (NM_000454): The protein encoded by this gene binds copper and zinc ions and is one of two isozymes responsible for destroying free superoxide radicals in the body. The encoded isozyme is a soluble cytoplasmic protein, acting as a homodimer to convert naturally-occuring but harmful superoxide radicals to molecular oxygen and hydrogen peroxide. The other isozyme is a mitochondrial protein. Mutations in this gene have been implicated as causes of familial amyotrophic lateral sclerosis. Rare transcript variants have been reported for this gene. [provided by RefSeq, Jul 2008]. Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Gene record to access additional publications. ##RefSeq-Attributes-START## Transcript_exon_combination_evidence :: BI517874.1, BI601911.1 [ECO:0000332] ##RefSeq-Attributes-END##
Transcript (Including UTRs):
Position: chr21:33,031,935-33,041,243 **Size:** 9,309 **Total Exon Count:** 5 **Strand:** +
Coding Region:
Position: chr21:33,032,083-33,040,891 **Size:** 8,809 **Coding Exon Count:** 5

Links.ra

name omim

shortLabel OMIM

tables refLink kgXref

idSql select refLink.omimId from kgXref,refLink where kgID = '%s' and kgXref.refseq = refLink.mrnaAcc and refLink.omimId != 0

url http://omim.org/%s

priority 10

hgGene Page

Genomic Sequence (chr21:33,031,935-33,041,243)			mRNA (may differ from genome)		Protein (154 aa)
Gene Sorter	Genome Browser	Protein FASTA	VisiGene	Table Schema	BioGPS
CGAP	Ensembl	Entrez Gene	ExonPrimer	GeneCards	GeneNetwork
Gepis Tissue	H-INV	HGNC	HPRD	Human Cortex Gene Expression	Jackson Lab
MOPED	OMIM	PubMed	Reactome	Stanford SOURCE	Treefam
UniProtKB	Wikipedia				

uniProt

Comments and Description Text from UniProtKB

ID: SODC_HUMAN
DESCRIPTION: RecName: Full=Superoxide dismutase [Cu-Zn]; EC=1.15.1.1; AltName: Full=Superoxide dismutase 1; Short=hSod1;
FUNCTION: Destroys radicals which are normally produced within the cells and which are toxic to biological systems.
CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
COFACTOR: Binds 1 copper ion per subunit.
COFACTOR: Binds 1 zinc ion per subunit.
SUBUNIT: Homodimer; non-disulfide linked. Homodimerization may take place via the ditryptophan cross-link at Trp-33. The pathogenic variants ALS1 Arg-38, Arg-47, Arg-86 and Ala-94 interact with RNF19A, whereas wild-type protein does not. The pathogenic variants ALS1 Arg-86 and Ala-94 interact with MARCH5, whereas wild-type protein does not.
INTERACTION: P26339:Chga (xeno); NbExp=5; IntAct=EBI-990792, EBI-990900; P16014:Chgb (xeno); NbExp=6; IntAct=EBI-990792, EBI-990820;
SUBCELLULAR LOCATION: Cytoplasm. Note=The pathogenic variants ALS1 Arg-86 and Ala-94 gradually aggregates and accumulates in mitochondria.
PTM: Unlike wild-type protein, the pathogenic variants ALS1 Arg- 38, Arg-47, Arg-86 and Ala-94 are polyubiquitinated by RNF19A leading to their proteasomal degradation. The pathogenic variants ALS1 Arg-86 and Ala-94 are ubiquitinated by MARCH5 leading to their proteasomal degradation.
PTM: The ditryptophan cross-link at Trp-33 is responsible for the non-disulfide-linked homodimerization. Such modification might only occur in extreme conditions and additional experimental evidence is required.
DISEASE: Defects in SOD1 are the cause of amyotrophic lateral sclerosis type 1 (ALS1) [MIM:105400]. ALS1 is a familial form of amyotrophic lateral sclerosis, a neurodegenerative disorder affecting upper and lower motor neurons and resulting in fatal paralysis. Sensory abnormalities are absent. Death usually occurs within 2 to 5 years. The etiology of amyotrophic lateral sclerosis is likely to be multifactorial, involving both genetic and environmental factors. The disease is inherited in 5-10% of cases leading to familial forms.
MISCELLANEOUS: The protein (both wild-type and ALS1 variants) has a tendency to form fibrillar aggregates in the absence of the intramolecular disulfide bond or of bound zinc ions. These aggregates may have cytotoxic effects. Zinc binding promotes dimerization and stabilizes the native form.
SIMILARITY: Belongs to the Cu-Zn superoxide dismutase family.
WEB RESOURCE: Name=Alsod; Note=ALS genetic mutations db; URL="http://alsod.iop.kcl.ac.uk/Als/";
WEB RESOURCE: Name=GeneReviews; URL="http://www.ncbi.nlm.nih.gov/sites/GeneTests/lab/gene/SOD1";
WEB RESOURCE: Name=NIEHS-SNPs; URL="http://egp.gs.washington.edu/data/sod1/";
WEB RESOURCE: Name=Wikipedia; Note=Superoxide dismutase entry; URL="http://en.wikipedia.org/wiki/Superoxide_dismutase";

gad
joined on uniProt id

Genetic Association Studies of Complex Diseases and Disorders

Genetic Association Database: SOD1
CDC HuGE Published Literature: SOD1
Positive Disease Associations: ALS/amyotrophic lateral sclerosis , amyotrophic lateral sclerosis cause novel protein interactions , familial amyotrophic lateral sclerosis , familial amyotrophic lateral sclerosis. , rapidly progressive familial amyotrophic lateral sclerosis , slowly progressive ALS
Related Studies:

1. **ALS/amyotrophic lateral sclerosis**
   Mancuso, M. et al. 2002, A screening for superoxide dismutase-1 D90A mutation in Italian patients with sporadic amyotrophic lateral sclerosis., Amyotrophic lateral sclerosis and other motor neuron disorders. 2002 Dec;3(4):215-8. [PubMed 12710511]
   Our results confirm that recessive D90A mutation is present in Italy and it is associated with the phenotype already described A screening for that mutation, easily made by RFLP, should be made in sporadic ALS patients, especially where clinical investigation indicates its presence.
2. **amyotrophic lateral sclerosis cause novel protein interactions**
   Kunst CB et al. 1997, Mutations in SOD1 associated with amyotrophic lateral sclerosis cause novel protein interactions., Nature genetics. 1997 Jan;15(1):91-4. [PubMed 8988176]
3. **familial amyotrophic lateral sclerosis**
   Morita M et al. 1996, A novel two-base mutation in the Cu/Zn superoxide dismutase gene associated with familial amyotrophic lateral sclerosis in Japan., Neuroscience letters. 1996 Feb;205(2):79-82. [PubMed 8907321]

more ... click here to view the complete list

Who makes this?

hgFixed.ctdSorted, kgXref.geneSymbol

Comparative Toxicogenomics Database (CTD)

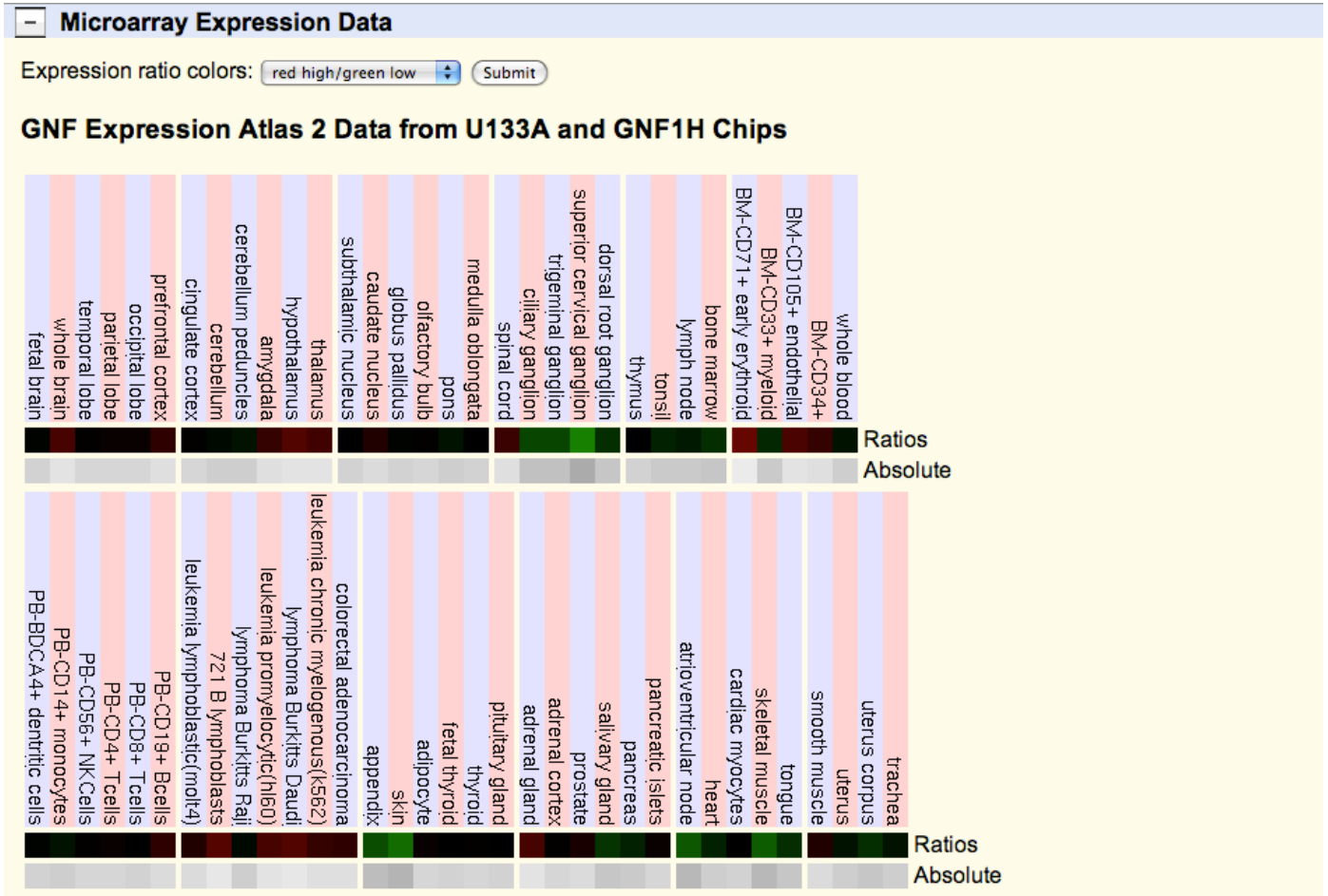
The following chemicals interact with this gene

- D010269 Paraquat
- D003300 Copper
- C467567 lenalidomide
- D019256 Cadmium Chloride
- D013792 Thalidomide
- D015032 Zinc
- D003314 Corn Oil
- D011078 Polychlorinated Biphenyls
- C017160 cypermethrin
- C017947 sodium arsenite

more ... click here to view the complete list

Who makes this?

GNF Atlas


Microarray Expression Data
Expression ratio colors: red high/green low
Submit
GNF Expression Atlas 2 Data from U133A and GNF1H Chips
whole blood
BM-CD34+
BM-CD105+ endothelial
BM-CD33+ myeloid
BM-CD71+ early erythroid
bone marrow
lymph node
tonsil
thymus
dorsal root ganglion
superior cervical ganglion
trigeminal ganglion
ciliary ganglion
spinal cord
medulla oblongata
pons
olfactory bulb
globus pallidus
caudate nucleus
subthalamic nucleus
thalamus
hypothalamus
amygdala
cerebellum peduncles
cerebellum
cingulate cortex
prefrontal cortex
occipital lobe
parietal lobe
temporal lobe
whole brain
fetal brain
Ratios
Absolute
trachea
uterus corpus
uterus
smooth muscle
tongue
skeletal muscle
heart
cardiac myocytes
atrioventricular node
pancreatic islets
pancreas
salivary gland
prostate
adrenal cortex
adrenal gland
pituitary gland
thyroid
fetal thyroid
adipocyte
skin
appendix
colorectal adenocarcinoma
leukemia chronic myelogenous(k562)
lymphoma Burkitts Daudi
leukemia promyelocytic(hl60)
lymphoma Burkitts Raji
721 B lymphoblasts
leukemia lymphoblastic(molt4)
PB-CD19+ Bcells
PB-CD8+ Tcells
PB-CD4+ Tcells
PB-CD56+ NKCells
PB-CD14+ monocytes
PB-BDCA4+ dentritic cells
Ratios
Absolute

foldUtr3, foldUtr5
indexed by UC id

mRNA Secondary Structure of 3' and 5' UTRs

Region	Fold Energy	Bases	Energy/Base	Display As		
5' UTR	-58.92	148	-0.398	Picture	PostScript	Text
3' UTR	-68.14	352	-0.194	Picture	PostScript	Text

The RNAfold program from the Vienna RNA Package is used to perform the secondary structure predictions and folding calculations. The estimated folding energy is in kcal/mol. The more negative the energy, the more secondary structure the RNA is likely to have.

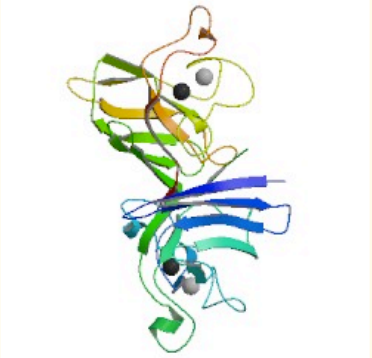
Protein Domain and Structure Information

InterPro Domains: Graphical view of domain structure
IPR024134 - SOD_Cu/Zn_/chaperones
IPR018152 - SOD_Cu/Zn_BS
IPR001424 - SOD_Cu_Zn_dom

Pfam Domains:
PF00080 - Copper/zinc superoxide dismutase (SODC)

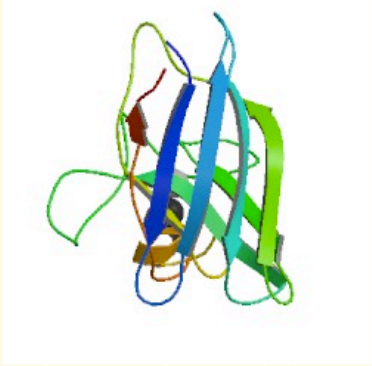
SCOP Domains:
49329 - Cu,Zn superoxide dismutase-like

Protein Data Bank (PDB) 3-D Structure



1AZV - X-ray Chimera LS-SNP

1BA9 - NMR Chimera LS-SNP

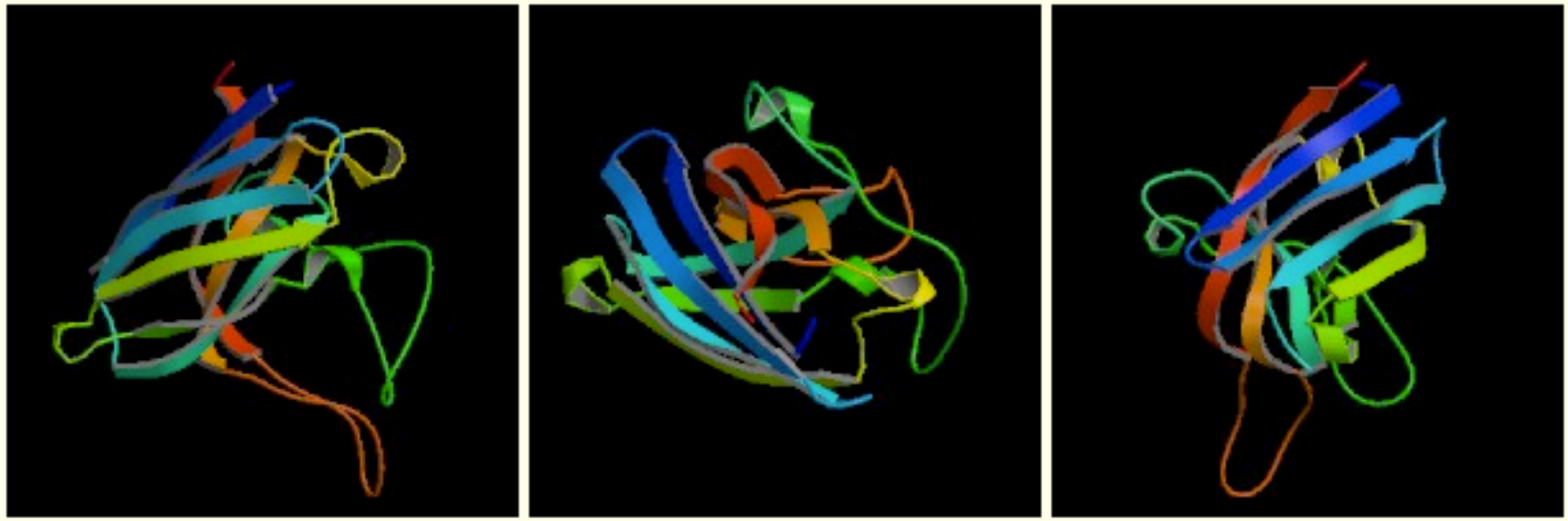


1DSW - NMR Chimera LS-SNP

To conserve bandwidth, only the images from the first 3 structures are shown.

1FUN - X-ray Chimera LS-SNP	1HL4 - X-ray Chimera LS-SNP	1HL5 - X-ray Chimera LS-SNP
1KMG - NMR Chimera LS-SNP	1L3N - NMR Chimera LS-SNP	1MFM - X-ray Chimera LS-SNP
1N18 - X-ray Chimera LS-SNP	1N19 - X-ray Chimera LS-SNP	1OEZ - X-ray Chimera LS-SNP
1OZT - X-ray Chimera LS-SNP	1OZU - X-ray Chimera LS-SNP	1P1V - X-ray Chimera LS-SNP

ModBase Predicted Comparative 3D Structure on P00441



Front Top Side

BlastTabs

Orthologous Genes in Other Species

Orthologies between human, mouse, and rat are computed by taking the best BLASTP hit, and filtering out non-syntenic hits. For more distant species reciprocal-best BLASTP hits are used. Note that the absence of an ortholog in the table below may reflect incomplete annotations in the other species rather than a true absence of the orthologous gene.

Mouse	Rat	Zebrafish	D. melanogaster	C. elegans	S. cerevisiae
Genome Browser	Genome Browser	Genome Browser	Genome Browser	Genome Browser	Genome Browser
Gene Details	Gene Details		Gene Details	Gene Details	Gene Details
Gene Sorter	Gene Sorter		Gene Sorter	Gene Sorter	Gene Sorter
Jackson Lab	RGD	Ensembl	FlyBase	WormBase	SGD
Protein Sequence	Protein Sequence	Protein Sequence	Protein Sequence	Protein Sequence	Protein Sequence
Alignment	Alignment	Alignment	Alignment	Alignment	Alignment

uniProt id to GO term

Gene Ontology (GO) Annotations with Structured Vocabulary

Molecular Function:

GO:0004784 superoxide dismutase activity
GO:0004785 copper, zinc superoxide dismutase activity
GO:0005507 copper ion binding
GO:0005515 protein binding
GO:0008270 zinc ion binding
GO:0016209 antioxidant activity
GO:0016491 oxidoreductase activity
GO:0030346 protein phosphatase 2B binding
GO:0042803 protein homodimerization activity
GO:0046872 metal ion binding
GO:0051087 chaperone binding

Who builds go database?

genbank

Descriptions from all associated GenBank mRNAs

DD328936 - Allele-Specific RNA Interference.
FU760796 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
DD328937 - Allele-Specific RNA Interference.
FU760801 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
JA482029 - Sequence 12 from Patent WO2011072091.
FU760795 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
FU760800 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
FU760802 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
AK312116 - Homo sapiens cDNA, FLJ92398, Homo sapiens superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult)) (SOD
BC001034 - Homo sapiens superoxide dismutase 1, soluble, mRNA (cDNA clone MGC:2325 IMAGE:3140145), complete cds.
X02317 - Human mRNA for Cu/Zn superoxide dismutase (SOD).
EF151142 - Homo sapiens superoxide dismutase 1 (SOD1) mRNA, complete cds.
EF143990 - Homo sapiens superoxide dismutase 1 (SOD1) mRNA, partial cds.
AB464254 - Synthetic construct DNA, clone: pF1KB8213, Homo sapiens SOD1 gene for superoxide dismutase 1, soluble, without stop cod
CR450355 - Homo sapiens full open reading frame cDNA clone RZPDo834A053D for gene SOD1, superoxide dismutase 1, soluble (amyo
without stopcodon.
BT006676 - Homo sapiens superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult)) mRNA, complete cds.
CU674512 - Synthetic construct Homo sapiens gateway clone IMAGE:100017938 5' read SOD1 mRNA.
AY450286 - Homo sapiens superoxide dismutase (SOD) mRNA, complete cds.
CR541742 - Homo sapiens full open reading frame cDNA clone RZPDo834E0529D for gene SOD1, superoxide dismutase 1, soluble (amy
stopcodon.
AY049787 - Homo sapiens soluble superoxide dismutase 1 (SOD1) gene, complete cds.
FU760798 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
E00383 - DNA coding of human superoxide dismutase.
E00882 - cDNA encoding human superoxide dismutase.
E06744 - cDNA fragment.
E06789 - cDNA encoding human super oxide dismutase(SOD).
E00834 - cDNA coding human Cu,Zn-superoxide dismutase.
E01552 - DNA encoding human superoxide dismutase.
BD174112 - Method of treating disease in association with decrease in the expression of AOP-1 gene or AOP-1 and remedies for the diseas
FU760799 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.
FU760797 - Methods and Compositions to Treat and Detect Misfolded-SOD1 Mediated Diseases.

Biochemical and Signaling Pathways

KEGG - Kyoto Encyclopedia of Genes and Genomes
hsa04146 - Peroxisome
hsa05014 - Amyotrophic lateral sclerosis (ALS)
hsa05016 - Huntington's disease
hsa05020 - Prion diseases

BioCyc Knowledge Library
DETOX1-PWY - superoxide radicals degradation

BioCarta from NCI Cancer Genome Anatomy Project
h_flumazenilPathway - Cardiac Protection Against ROS
h_freePathway - Free Radical Induced Apoptosis
h_longevityPathway - The IGF-1 Receptor and Longevity

Reactome (by CSHL, EBI, and GO)

Protein P00441 (Reactome details) participates in the following event(s):

482772 Release of platelet cytosolic components
114608 Platelet degranulation
76005 Response to elevated platelet cytosolic Ca2+
76002 Platelet activation, signaling and aggregation
109582 Hemostasis

Other Names for This Gene

Alternate Gene Symbols: A6NHJ0, D3DSE4, NM_000454, NP_000445, P00
UCSC ID: uc002ypa.3
RefSeq Accession: NM_000454
Protein: P00441 (aka SODC_HUMAN)
CCDS: CCDS33536.1

GeneReviews for This Gene

GeneReview(s) and GeneTest disease(s) related to gene SOD1:
als-overview (Amyotrophic Lateral Sclerosis)

kgTxInfo

Gene Model Information

category:	coding	**nonsense-mediated-decay:**	no	**RNA accession:**	NM_000454.4
exon count:	5	**CDS single in 3' UTR:**	no	**RNA size:**	981
ORF size:	465	**CDS single in intron:**	no	**Alignment % ID:**	100.00
txCdsPredict score:	1105.00	**frame shift in genome:**	no	**% Coverage:**	98.37
has start codon:	yes	**stop codon in genome:**	no	**# of Alignments:**	1
has end codon:	yes	**retained intron:**	no	**# AT/AC introns**	0
selenocysteine:	no	**end bleed into intron:**	0	**# strange splices:**	0

Click here for a detailed description of the fields of the table above.

kgTxInfo

Field	Type	Null	Key	Default	Extra
name	varchar(255)	NO	MUL	NULL	
category	varchar(255)	NO		NULL	
sourceAcc	varchar(255)	NO		NULL	
isRefSeq	tinyint(3) unsigned	NO		NULL	
sourceSize	int(11)	NO		NULL	
aliCoverage	double	NO		NULL	
aliIdRatio	double	NO		NULL	
genoMapCount	int(11)	NO		NULL	
exonCount	int(11)	NO		NULL	
orfSize	int(11)	NO		NULL	
cdsScore	double	NO		NULL	
startComplete	tinyint(3) unsigned	NO		NULL	
endComplete	tinyint(3) unsigned	NO		NULL	
nonsenseMediatedDecay	tinyint(3) unsigned	NO		NULL	
retainedIntron	tinyint(3) unsigned	NO		NULL	
bleedIntoIntron	int(11)	NO		NULL	
strangeSplice	int(11)	NO		NULL	
atacIntrons	int(11)	NO		NULL	
cdsSingleInIntron	tinyint(3) unsigned	NO		NULL	
cdsSingleInUtr3	tinyint(3) unsigned	NO		NULL	
selenocysteine	tinyint(3) unsigned	NO		NULL	
genomicFrameShift	tinyint(3) unsigned	NO		NULL	
genomicStop	tinyint(3) unsigned	NO		NULL	

Thanks Everyone!