

refTSS: A reference dataset for human and mouse transcription start sites

User manual

About refTSS

- Transcription starts at genomic positions called transcription start sites (TSSs), producing RNAs, and is mainly regulated by genomic elements and transcription factors binding around these TSSs.
- refTSS is a comprehensive reference set that integrates all known TSSs. refTSS is a reference dataset of transcription start sites (refTSS) for the human and mouse genomes.
- The refTSS dataset consists of genomic coordinates of TSS peaks and their gene annotations.

The content of refTSS user manual

refTSS web interface was built using semantic Wikimedia a free, open-source extension to [MediaWiki](#) *.

The user manual consists of the following:

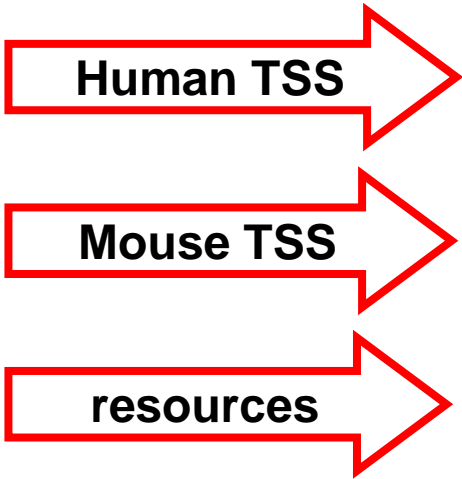
- 1- Overview of the main page (landing page).
 - 1-1 The data files
 - 1-2 refTSS UCSC genome browser track hub
 - 1-3 Contact information and support
- 2- The human TSS sub-menu (this is identical to the mouse)
 - 2-1 TSS query form
 - 2-2 Entrez gene query form human
- 3- The data sources for download

* https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki

refTSS web interface: main page

refTSS: reference transcription start sites

Search Go Search



Navigation

- Main Page
- Human**
 - TSSs
 - gene
- Mouse**
 - TSSs
 - gene
- Resources**
 - Data source

Main Page

Contents [hide]

- 1 refTSS: reference transcription start site
 - 1.1 Data files
 - 1.2 UCSC genome browser Track Hubs
 - 1.3 Contact

refTSS: reference transcription start site

The refTSS is an annotated reference dataset for transcriptional start sites (TSS) in human and mouse. The dataset is generated by collecting, reprocessing and assembling various public resources.

Data files

The raw refTSS data files are available at [the data download page](#).

UCSC genome browser Track Hubs

The hub.txt for UCSC genome browser "Track Hubs" is available at <http://refTSS.cst.riken.jp/trackhub/hub.txt>.

Contact

If you have any questions or comments, please contact to refTSS-help@riken.jp. ("[at](mailto:refTSS-help@riken.jp)" should be replaced with "@")

Data files

- The data files provide links to the latest version of refTSS and old versions.

refTSS: reference transcription start sites

Search Go Search

Navigation

- Main Page

Human

- TSSs
- gene

Mouse

- TSSs
- gene

Resources

- Data source

Main Page

Contents [hide]

- 1 refTSS: reference transcription start site
 - 1.1 Data files
 - 1.2 UCSC genome browser Track Hubs
 - 1.3 Contact

refTSS: reference transcription start site

The refTSS is an annotated reference dataset for transcriptional start sites (TSS) in human and mouse. The dataset is generated by collecting, reprocessing and assembling various public resources.

Data files

The raw refTSS data files are available at [the data download page](#)

UCSC genome browser Track Hubs

The hub.txt for UCSC genome browser "Track Hubs" is available at <http://refTSS.cst.riken.jp/trackhub/hub.txt>

Contact

If you have any questions or comments, please contact to [refTSS-help<at>riken.jp](mailto:refTSS-help@riken.jp) . ("[at](mailto:refTSS-help@riken.jp)" should be replaced with "@")

UCSC genome browser track hub

- The UCSC genome browser track hub enable the user to browse refTSS track hub on UCSC genome browser.
- Hub.txt files is located under <http://refTSS.clst.riken.jp/trackhub/hub.txt>
- User paste the hub.txt file ulr

The screenshot shows the refTSS website interface. At the top, there is a dark blue header with the text "refTSS: reference transcription start sites" and a search bar with "Search" and "Go" buttons. On the left side, there is a navigation menu with sections: "Navigation" (Main Page), "Human" (TSSs, gene), "Mouse" (TSSs, gene), and "Resources" (Data source). The main content area is titled "Main Page" and contains a "Contents [hide]" box with a list of links: "1 refTSS: reference transcription start site", "1.1 Data files", "1.2 UCSC genome browser Track Hubs", and "1.3 Contact". Below this, there is a section titled "refTSS: reference transcription start site" with a description: "The refTSS is an annotated reference dataset for transcriptional start sites (TSS) in human and mouse. The dataset is generated by collecting, reprocessing and assembling various public resources." This is followed by a "Data files" section stating "The raw refTSS data files are available at the data download page". A red box highlights the "UCSC genome browser Track Hubs" section, which states "The hub.txt for UCSC genome browser 'Track Hubs' is available at <http://refTSS.clst.riken.jp/trackhub/hub.txt>". The final section is "Contact", which provides an email address: "If you have any questions or comments, please contact to refTSS-help<at>riken.jp . ("<at>" should be replaced with "@")".

Display refTSS UCSC track hub

Track Data Hubs

Track data hubs are collections of external tracks that can be imported into the UCSC Genome Browser. Hub tracks show up under the hub's own blue label bar on the main browser page, as well as on the configure page. For more information, including where to [host](#) your track hub, see the [User's Guide](#). To import a public hub click its "Connect" button below.

NOTE: Because Track Hubs are created and maintained by external sources, UCSC is not responsible for their content.

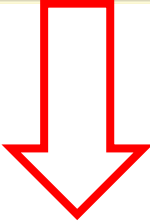
Public Hubs My Hubs

URL: Add Hub

No Unlisted Track Hubs

Contact us to add a public hub.

Paste hub.txt file url



Track Data Hubs

Track data hubs are collections of external tracks that can be imported into the UCSC Genome Browser. Hub tracks show up under the hub's own blue label bar on the main browser page, as well as on the configure page. For more information, including where to [host](#) your track hub, see the [User's Guide](#). To import a public hub click its "Connect" button below.

NOTE: Because Track Hubs are created and maintained by external sources, UCSC is not responsible for their content.

Public Hubs My Hubs

Display	Hub Name	Description	Assemblies
<input type="button" value="Disconnect"/>	refTSS	refTSS	hg38, mm10

Contact us to add a public hub.

Add the hub

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr1:11,102,837-11,267,747 164,911 bp. enter position, gene symbol, HGVS or search terms go

chr1 (p36.22) 105111 10112 10441 105144

TSS peaks
OH1M Alleles
DNase Clusters
Common SNPs (151)
Alt Haplotypes
obTSS
DR000914
EPDneu
FANTOM5
RANPAGE

FANTOM5 DPI peak, robust set
OH1M Allelic Variants
DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types)
Single Nucleotide Use Polymorphisms (SNUP 151) Found in >= 15 of Samples
Reference Assembly Alternate Haplotype Sequence Alignments

refTSS
obTSS
DR000914
EPDneu
FANTOM5
RANPAGE

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

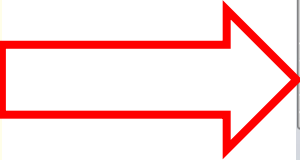
track search default tracks default order hide all add custom tracks track hubs configure multi-region reverse resize refresh collapse all expand all

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

FANTOM5 disconnect refresh

refTSS sources... disconnect refresh

full show



refTSS track hub will be added

The TSS query form for human

- For each genome (human and mouse), the system provide two interface, [1] TSS query menu and [2] Gene query menu.
- The TSS query menu enable the user to search for TSS by any of the following argument:
- **1** TSS ID, **2** Gene symbol and **3** Gene name

The screenshot displays the 'refTSS: reference transcription start sites' web application. At the top, there is a search bar with 'Search' and 'Go' buttons, and a 'Special page' link. The main content area is titled 'Run query: TSSQueryFormHuman'. On the left, a navigation menu includes 'Main Page', 'Human' (with sub-items 'TSSs' and 'gene'), 'Mouse' (with sub-items 'TSSs' and 'gene'), and 'Resources' (with 'Data source'). The main form area contains a 'Search items(limit 2000 lines)' section with three input fields: 'TSS ID:' (marked with a circled '1'), 'Gene Symbol:' (marked with a circled '2'), and 'Gene Name:' (marked with a circled '3'). A 'Run query' button is located below these fields. The footer contains links for 'Privacy policy', 'About referenceTSS', and 'Disclaimers'.

Search by gene symbol

- Search for Gapdh gene will retrieve all TSS, that found to be associated with Gapdh.
- Alternatively, use can search by gene Name

refTSS: reference transcription start sites [Special page](#)

Navigation

- Main Page
- Human
 - TSSs
 - gene
- Mouse
 - TSSs
 - gene
- Resources
 - Data source

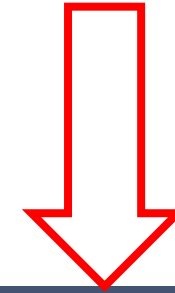
Run query: TSSQueryFormHuman

Search items(limit 2000 lines)

TSS ID:

Gene Symbol:

Gene Name:



refTSS: reference transcription start sites [Special page](#)

Navigation

- Main Page
- Human
 - TSSs
 - gene
- Mouse
 - TSSs
 - gene
- Resources
 - Data source

Run query: TSSQueryFormHuman

Show 25 entries Search:

TSS ID	Chr	Gene Symbol	Start	End	Strand
TSS:Hg 123688.1	chr12	GAPDH	6,534,510	6,534,527	+
TSS:Hg 123689.1	chr12	GAPDH	6,535,224	6,535,229	+
TSS:Hg 123690.1	chr12	GAPDH	6,535,231	6,535,273	+
TSS:Hg 123691.1	chr12	GAPDH	6,535,277	6,535,303	+
TSS:Hg 123692.1	chr12	GAPDH	6,535,307	6,535,322	+
TSS:Hg 123693.1	chr12	GAPDH	6,535,344	6,535,345	+
TSS:Hg 123694.1	chr12	GAPDH	6,535,357	6,535,410	+
TSS:Hg 123695.1	chr12	GAPDH	6,536,470	6,536,511	+
TSS:Hg 123696.1	chr12	GAPDH	6,536,680	6,536,698	+
TSS:Hg 123697.1	chr12	GAPDH	6,536,938	6,536,961	+
TSS:Hg 123698.1	chr12	GAPDH	6,537,128	6,537,169	+
TSS:Hg 123700.1	chr12	GAPDH	6,537,632	6,537,661	+
TSS:Hg 123701.1	chr12	GAPDH	6,537,752	6,537,775	+
TSS:Hg 123702.1	chr12	GAPDH	6,537,781	6,537,800	+

TSS detailed information

- Clicking any of the TSS ID, detailed information about the TSS are provided.

refTSS: reference transcription start sites

Search Go [Special page](#)

Navigation

- Main Page
- Human
 - TSSs
 - gene
- Mouse
 - TSSs
 - gene
- Resources
 - Data source

Run query: TSSQueryFormHuman

TSS ID	Chr	Gene Symbol	Start	End	Strand
TSS:Hg 110642.1	chr10	GAPDHP28	91,667,518	91,667,520	-
TSS:Hg 123688.1	chr12	GAPDH	6,534,510	6,534,527	+
TSS:Hg 123689.1	chr12	GAPDH	6,535,224	6,535,229	+
TSS:Hg 123690.1	chr12	GAPDH	6,535,231	6,535,273	+
TSS:Hg 123691.1	chr12	GAPDH	6,535,277	6,535,303	+
TSS:Hg 123692.1	chr12	GAPDH	6,535,307	6,535,322	+
TSS:Hg 123693.1	chr12	GAPDH	6,535,344	6,535,345	+
TSS:Hg 123694.1	chr12	GAPDH	6,535,357	6,535,410	+
TSS:Hg 123695.1	chr12	GAPDH	6,536,470	6,536,511	+
TSS:Hg 123696.1	chr12	GAPDH	6,536,680	6,536,698	+
TSS:Ho 123697.1	chr12	GAPDH	6,536,938	6,536,961	+

TSS:Ho 110642.1

Basic information about the TSS

TSS ID: hg_110642.1 Start: 91667518
Species: human End: 91667520
Chr: chr10 Strand: -
TSS Classification: no TSS Score: 1

Gene and protein annotation

Transcript_name: ENST00000455134.1 Gene_name: glyceraldehyde 3 phosphate
Distance: -262 Gene_symbol: GAPDHP28
GeneID: pseudogene 28
HGNC/MGI_ID: HGNC:37780 Gene_synonyms:
UniProt_ID: Gene_source: HGNC:37780

Transcript annotation

Transcript_name: ENST00000455134.1
Distance: -262
Num_trans: 1
All_trans: ENST00000455134.1:262

TATA-Box annotation

Distance to TSS: -34418
Nearest PromoterID: NM_005398
Entrez ID: 5507
Nearest Unigene: Hs_303090
Nearest Refseq: NM_005398
Nearest Ensembl: ENSG00000119938
Gene Name: glyceraldehyde 3 phosphate dehydrogenase pseudogene 28
Gene Alias: PPP1R5
Gene Description: protein phosphatase 1 regulatory subunit 3C
Gene Type: protein-coding
CpG%: 0.005
GC%: 0.442786
TATA-Box(TBP)/Promoter/Homer Distance From: *-90(CACCATTAAAGC,-,0.00)*
TSS(sequence,strand,conservation):

Entrez gene query form for human

- The gene query menu enable the user to search for TSS by any of the following argument:
- **1** Gene symbol and **2** Gene name
- Search GAPDH gene, will result in a list of all related TSS.

The image shows three sequential screenshots of the Entrez Gene query form for human, illustrating the search process for TSSs of the GAPDH gene.

Top Screenshot: Query Form
The page title is "refTSS: reference transcription start sites". The search form is titled "Run query: EntrezGeneQueryFormHuman". It has a search bar with "Search" and "Go" buttons. Below the search bar, there are two input fields: "GENE ID:" (labeled with a circled "1") and "Gene Symbol:" (labeled with a circled "2"). A "Run query" button is located below the form.

Middle Screenshot: Search Results
The page title is "refTSS: reference transcription start sites". The search form is titled "Run query: EntrezGeneQueryFormHuman". Below the search bar, there are buttons for "Copy", "Excel", "CSV", and "PDF", and a "Show 25 entries" dropdown. The search results are displayed in a table with two columns: "GENE ID" and "Gene Symbol".

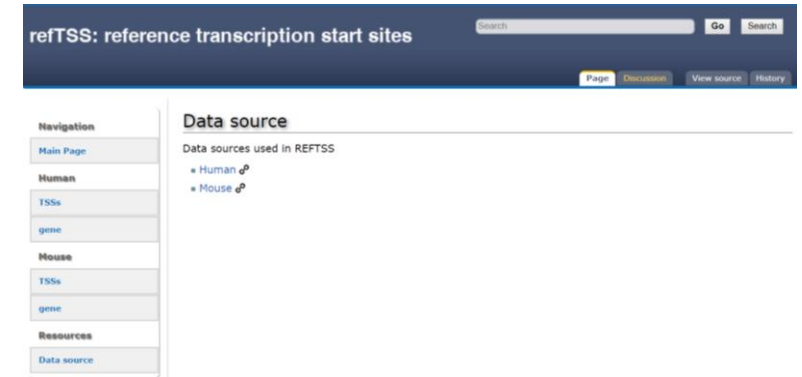
GENE ID	Gene Symbol
EntrezGene:2597	GAPDH
EntrezGene:2616	GAPDHP1
EntrezGene:54050	GAPDHP14
EntrezGene:642317	GAPDHP15
EntrezGene:387491	GAPDHP16
EntrezGene:360008	GAPDHP17
EntrezGene:140028	GAPDHP19
EntrezGene:170516	GAPDHP?

Bottom Screenshot: Detailed View
The page title is "refTSS: reference transcription start sites". The search form is titled "EntrezGene: 2597". Below the search bar, there are buttons for "EntrezGene", "PubMed", "View source", and "History". The search results are displayed in a table with one column: "Related TSS IDs".

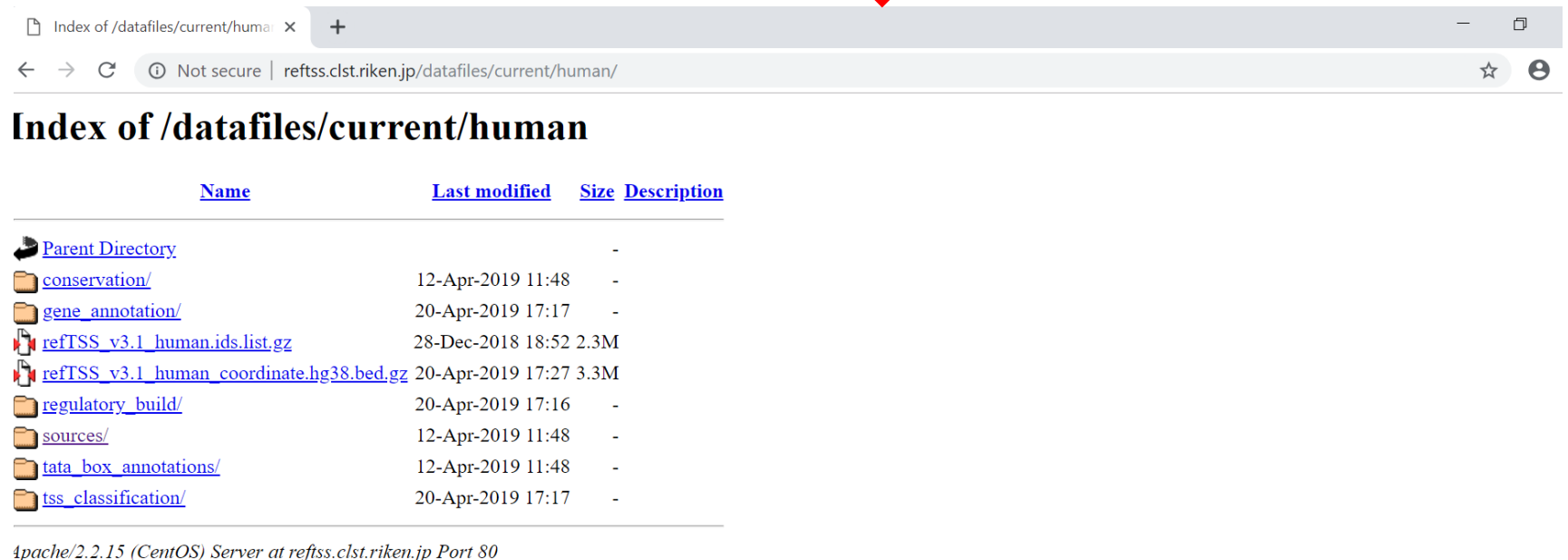
Related TSS IDs
Hg 123688.1
Hg 123689.1
Hg 123690.1
Hg 123691.1
Hg 123692.1
Hg 123693.1
Hg 123694.1
Hg 123695.1
Hg 123696.1
Hg 123697.1
Hg 123698.1
Hg 123700.1
Hg 123701.1
Hg 123702.1
Hg 123703.1
Hg 123704.1
Hg 123705.1

Data sources for download

- The data sources link provide access to all the datasets used to construct refTSS and the resulting annotation and QC.



The screenshot shows the refTSS website interface. At the top, there is a search bar and navigation links. The 'Data source' section is highlighted, showing 'Data sources used in REFTSS' with links for 'Human' and 'Mouse'. A red arrow points from this section down to the browser screenshot below.



The screenshot shows a web browser window displaying the 'Index of /datafiles/current/human' directory. The browser address bar shows the URL 'refTSS.clst.riken.jp/datafiles/current/human/'. The page title is 'Index of /datafiles/current/human'. Below the title is a table listing files and directories with columns for Name, Last modified, Size, and Description.

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
Parent Directory		-	
conservation/	12-Apr-2019 11:48	-	
gene_annotation/	20-Apr-2019 17:17	-	
refTSS_v3.1_human.ids.list.gz	28-Dec-2018 18:52	2.3M	
refTSS_v3.1_human_coordinate.hg38.bed.gz	20-Apr-2019 17:27	3.3M	
regulatory_build/	20-Apr-2019 17:16	-	
sources/	12-Apr-2019 11:48	-	
tata_box_annotations/	12-Apr-2019 11:48	-	
tss_classification/	20-Apr-2019 17:17	-	

Apache/2.2.15 (CentOS) Server at refTSS.clst.riken.jp Port 80