

Web-based tools for bioinformatics

BCMB bootcamp

bioinformatics tools have two major roles

- integrate public data and resources into new research
- enable analyses that are too large or complex to be done by hand

Two major approaches to bioinformatics

- command line: programs, data accessible locally or on a server, commands written by hand in a terminal-style application
 - easy to script complicated workflows
 - flexible! use any program you can install
 - requires familiarity with command line tools
- web-based: access databases programmatically via GUIs on web pages
 - no special scripting required
 - limited to databases and tools that have been embedded in web pages
 - larger scale projects may be more difficult to manage reproducibly

Outline for today

- GenBank: biggest sequence database
- Ensembl/BioMart: extremely useful tools for fetching data
- UCSC Genome Browser: web server for viewing and fetching genomic data
- Galaxy: web-based platform for bioinformatics analysis

GenBank

- NIH genetic sequence database
- part of the International Nucleotide Sequence Database Collaboration: DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI
- number of bases has doubled every 18 months since 1982!

GenBank

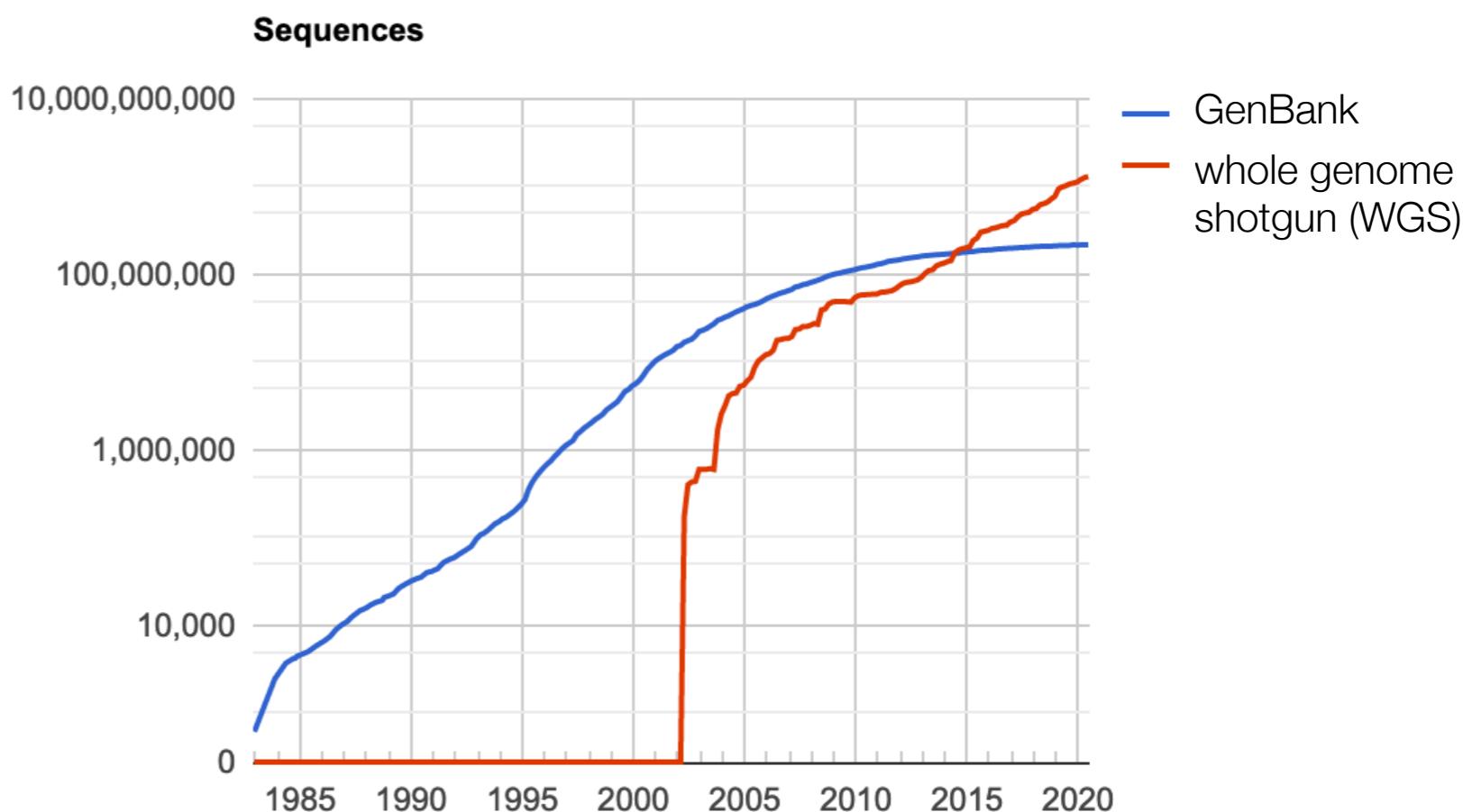
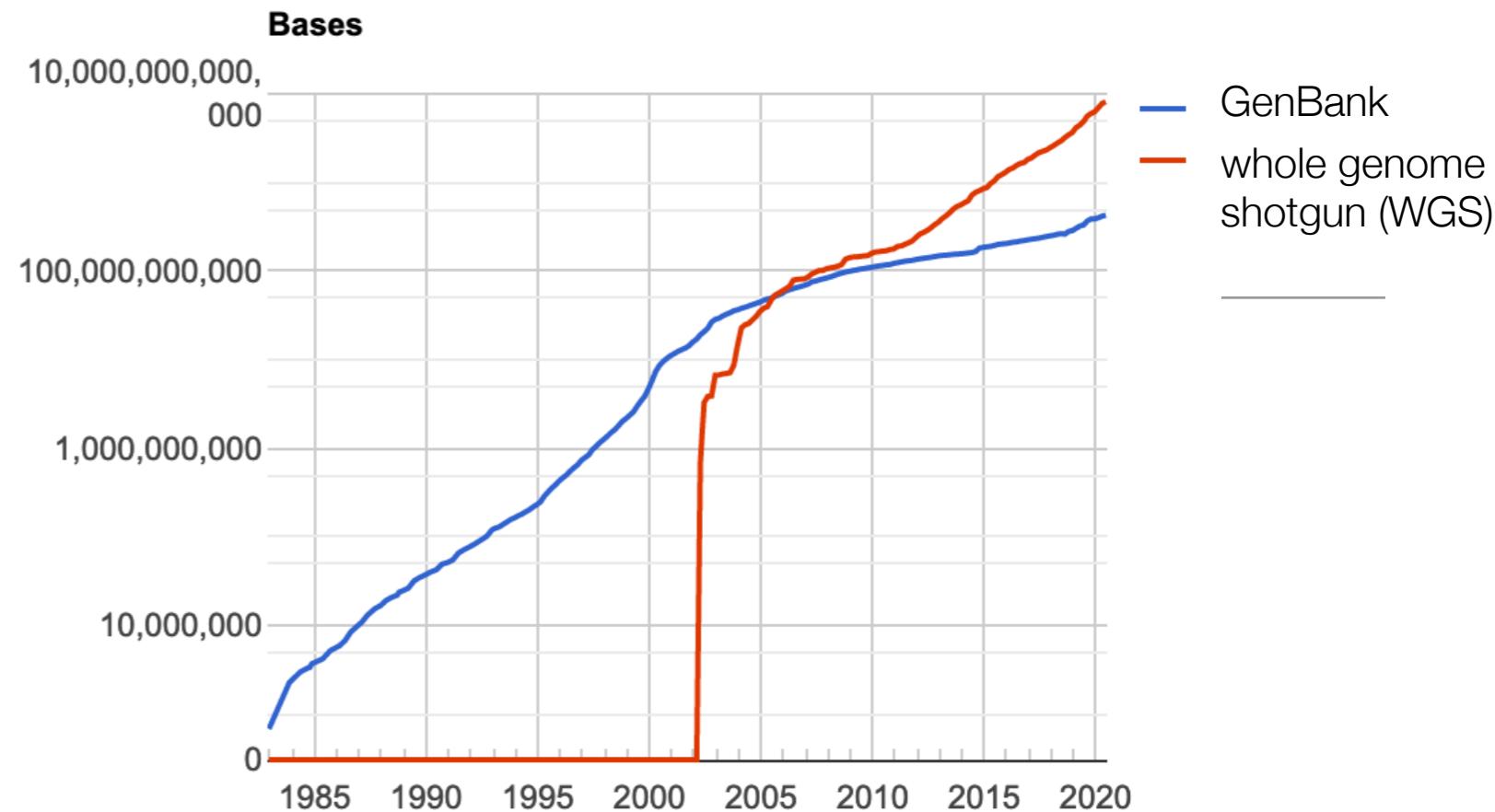
release 238 June 2020

GenBank:

427,823,258,901 bases
217,122,233 sequences

WGS:

8,114,046,262,158 bases
1,302,852,615 sequences



accessing GenBank

- NCBI website via Entrez
- ftp downloads
- search and download sequences from command line using NCBI e-utilities
- BLAST

www.ncbi.nlm.nih.gov/nuccore

search for TP53



NCBI Resources ▾ How To ▾

swheelan My NCBI Sign Out

Nucleotide Nucleotide ▾ | Search Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.



Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Using Nucleotide

[Quick Start Guide](#)

[FAQ](#)

[Help](#)

[GenBank FTP](#)

[RefSeq FTP](#)

Nucleotide Tools

[Submit to GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[BLAST](#)

[Batch Entrez](#)

Other Resources

[GenBank Home](#)

[RefSeq Home](#)

[Gene Home](#)

[SRA Home](#)

[INSDC](#)

Species

Animals (10,666)

Plants (127)

Fungi (84)

Protists (63)

Bacteria (73)

Archaea (22)

Viruses (31)

Customize ...

Molecule types

genomic

DNA/RNA (5,090)

mRNA (5,848)

Customize ...

Source databasesINSDC (GenBank)
(4,811)

RefSeq (6,868)

Customize ...

Sequence Type

Nucleotide (10,487)

EST (1,200)

GSS (3)

Genetic compartments

Mitochondrion (1)

Sequence length

Custom range...

Release date

Custom range...

Revision date

Custom range...

Summary ▾ 20 per page ▾ Sort by Default order ▾

Send to: ▾

Filters: [Manage Filters](#)**Results by taxon****Top Organisms [Tree]**[Homo sapiens \(3313\)](#)[synthetic construct \(536\)](#)[Mus musculus \(106\)](#)[Rattus norvegicus \(90\)](#)[Pan troglodytes \(73\)](#)[All other taxa \(7572\)](#)[More...](#)**Find related data**Database: [Select](#)[Find items](#)**Search details**

TP53 [All Fields]

[Search](#)[See more...](#)**Recent activity**[Turn Off](#) [Clear](#) [TP53 \(11690\)](#)

Nucleotide

GENE

Was this helpful?

**TP53 – tumor protein p53**[Homo sapiens \(human\)](#)

Also known as: BCC7, BMFS5, LFS1, P53, TRP53

GeneID: 7157

[RefSeq transcripts \(15\)](#)[RefSeq proteins \(15\)](#)[RefSeqGene \(1\)](#)[PubMed \(9,868\)](#)[Orthologs](#)[Genome Browser](#)[BLAST](#)[Download](#)**RefSeq Sequences****Items: 1 to 20 of 11690**

<< First < Prev Page 1 of 585 Next > Last >>

 [Homo sapiens TP53 \(TP53\) gene, exon 5 and partial cds](#)

1. 123 bp linear DNA

Accession: MH011443.1 GI: 1486783306

[Protein](#) [Taxonomy](#)[GenBank](#) [FASTA](#) [Graphics](#) [Homo sapiens isolate 88F TP53 \(TP53\) gene, exon 8 and partial cds](#)

2. 57 bp linear DNA

Accession: MH366483.1 GI: 1472901973

[Protein](#) [Taxonomy](#)

Complex Analysis of Retroposed Genes' Contribution to Human

ORF Capture-Seq as a versatile method for targeted identification

Reference Sequence database (RefSeq)

- a collection of taxonomically diverse, non-redundant and richly annotated sequences representing naturally occurring molecules of DNA, RNA, and protein. Included are sequences from plasmids, organelles, viruses, archaea, bacteria, and eukaryotes.
- Each RefSeq is constructed wholly from sequence data submitted to the International Nucleotide Sequence Database Collaboration (INSDC).
- Similar to a review article, a RefSeq is a synthesis of information integrated across multiple sources at a given time. RefSeqs provide a foundation for uniting sequence data with genetic and functional information.
- They are generated to provide reference standards for multiple purposes ranging from genome annotation to reporting locations of sequence variation in medical records.

RefSeq accessions

Accession prefix	Molecule type	Comment
AC_	Genomic	Complete genomic molecule, usually alternate assembly
NC_	Genomic	Complete genomic molecule, usually reference assembly
NG_	Genomic	Incomplete genomic region
NT_	Genomic	Contig or scaffold, clone-based or WG
NW_	Genomic	Contig or scaffold, primarily WG
NZ	Genomic	Complete genomes and unfinished WGS data
NM_	mRNA	Protein-coding transcripts (usually curated)
NR_	RNA	Non-protein-coding transcripts
XM	mRNA	Predicted model protein-coding transcript
XR	RNA	Predicted model non-protein-coding transcript
AP_	Protein	Annotated on AC_ alternate assembly
NP_	Protein	Associated with an NM_ or NC_ accession
YP	Protein	Annotated on genomic molecules without an instantiated transcript record
XP	Protein	Predicted model, associated with an XM_ accession
WP_	Protein	Non-redundant across multiple strains and species

search for TP53

The screenshot shows the NCBI RefSeq homepage. At the top, there's a blue header bar with the NCBI logo, 'Resources' dropdown, 'How To' dropdown, and user account links ('swheelan', 'My NCBI', 'Sign Out'). Below the header is a navigation bar with 'RefSeq' and a dropdown menu set to 'RefSeq'. To the right is a search bar with a 'Search' button. A large black arrow points downwards from the word 'search' in the main text above to the search bar on the page. A red banner at the top of the main content area contains a warning icon, COVID-19 information, and a link to SARS-CoV-2 literature.

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

The main content area features a dark blue banner with the text 'RefSeq: NCBI Reference Sequence Database' in white. Below the banner, a blue background image shows a DNA helix and chemical structures. To the right of the image, the text reads: 'A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.'

Using RefSeq

[About RefSeq](#)

[Human Reference Genome](#)

[Prokaryotic RefSeq Genomes](#)

[FAQ](#)

[NCBI Handbook](#)

[Factsheet](#)

RefSeq Access

[Human Genome Resources and Download](#)

[RefSeq FTP](#)

[RefSeq_genomes FTP](#)

[New RefSeq_genomic \(last 30 days\)](#)

[New RefSeq_transcripts \(last 30 days\)](#)

[New RefSeq_proteins \(last 30 days\)](#)

[Searching for RefSeq records \(Queries\)](#)

RefSeq projects

[Consensus CDS \(CCDS\)](#)

[RefSeq Functional Elements](#)

[RefSeqGene](#)

[Targeted Loci](#)

[Virus Variation](#)

[RefSeq Select](#)

[MANE](#)

Announcements

May 12, 2020

RefSeq Release 200 is available for FTP

This release includes:

Proteins: 171,643,720

Related Links

[Assembly](#)

[Gene](#)

[Genome](#)

Feedback & Credits

[Publications and Citing RefSeq](#)

[Contact RefSeq Help Desk](#)

[Contact CCDS Help Desk](#)

[Animals \(6,760\)](#)[Plants \(74\)](#)[Fungi \(17\)](#)[Protists \(17\)](#)[Customize ...](#)[Molecule](#)[types](#)[genomic](#)[DNA/RNA \(2,450\)](#)[mRNA \(4,238\)](#)[Customize ...](#)[Source](#)[databases](#)[RefSeq \(6,868\)](#)[Customize ...](#)[Sequence](#)[Type](#)[Nucleotide \(6,868\)](#)[Sequence](#)[length](#)[Custom range...](#)[Release date](#)[Custom range...](#)[Revision date](#)[Custom range...](#)[Clear all](#)[Show additional filters](#)

GENE

Was this helpful?



[TP53 orthologs from vertebrates](#)

tumor protein p53

[How are orthologs calculated?](#)

[Genes with similar protein architectures](#)

[How are similar genes calculated?](#)

Items: 1 to 20 of 6868

<< First < Prev Page 1 of 344 Next > Last >>

- [Canis lupus familiaris tumor protein p53 \(TP53\), mRNA](#)

1. 1,174 bp linear mRNA

Accession: NM_001003210.1 GI: 50978973

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

- [Homo sapiens tumor protein p53 \(TP53\), RefSeqGene \(LRG_321\) on](#)

2. [chromosome 17](#)

32,772 bp linear DNA

Accession: NG_017013.2 GI: 383209646

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

- [Bos taurus tumor protein p53 \(TP53\), mRNA](#)

Results by taxon

Top Organisms [Tree]

[Homo sapiens \(399\)](#)[Mus musculus \(69\)](#)[Rattus norvegicus \(64\)](#)[Danio rerio \(60\)](#)[Aotus nancymaae \(43\)](#)[All other taxa \(6233\)](#)[More...](#)

Find related data

Database: [Select](#)[Find items](#)

Search details

TP53[All Fields] AND
srcdb_refseq[PROP]

[Search](#)[See more...](#)

Recent activity

[Turn Off](#) [Clear](#)

TP53 AND srcdb_refseq[PROP] (6868) Nucleotide

TP53 (11690) Nucleotide

Complex Analysis of Retroposed Genes' Contribution to Human

www.ncbi.nlm.nih.gov/gene search for TP53

The screenshot shows the NCBI Gene search interface. At the top, there's a blue header bar with the NCBI logo, 'Resources' dropdown, 'How To' dropdown, user 'swheelan', 'My NCBI', and 'Sign Out'. Below the header is a search bar with 'Gene' selected, a dropdown set to 'TP53', and a 'Search' button. An arrow points down from the search term 'TP53' to a red banner at the bottom of the search results page.

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene

[Gene Quick Start](#)

[FAQ](#)

[Download/FTP](#)

[RefSeq Mailing List](#)

[Gene News](#)

[Factsheet](#)

Gene Tools

[Submit GeneRIFs](#)

[Submit Correction](#)

[Statistics](#)

[BLAST](#)

[Genome Workbench](#)

[Splign](#)

Other Resources

[OMIM](#)

[RefSeq](#)

[RefSeqGene](#)

[Protein Clusters](#)

Representative queries

[Find genes by...](#)

[Search text](#)

Search results

Items: 1 to 20 of 5574

<< First < Prev Page 1 of 279 Next > Last >>

 See also 124 discontinued or replaced items.

Search

See more...

Name/Gene ID	Description	Location	Aliases	MIN
<input type="checkbox"/> TP53 ID: 7157	tumor protein p53 [<i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (7668402..7687550, complement)	BCC7, BMFS5, LFS1, P53, TRP53	191
<input type="checkbox"/> Tp53 ID: 24842	tumor protein p53 [<i>Rattus norvegicus</i> (Norway rat)]	Chromosome 10, NC_005109.4 (56186299..56198449)	Trp53, p53	
<input type="checkbox"/> tp53 ID: 30590	tumor protein p53 [<i>Danio rerio</i> (zebrafish)]	Chromosome 5, NC_007116.7 (24086227..24097807)	brp53, drp53, etID22686.5, fb40d06, p53, wu:fb40d06, zgc:111919	
<input type="checkbox"/> TP53 ID: 403869	tumor protein p53 [<i>Canis lupus familiaris</i> (dog)]	Chromosome 5, NC_006587.3 (32561406..32565149, complement)	P53	
<input type="checkbox"/> TP53 ID: 397276	tumor protein p53 [<i>Sus scrofa</i> (pig)]	Chromosome 12, NC_010454.4 (52939643..52953786, complement)	P53	
<input type="checkbox"/> TP53 ID: 281542	tumor protein p53 [<i>Bos taurus</i> (cattle)]	Chromosome 19, NC_037346.1 (27376071..27388474, complement)		
<input type="checkbox"/> TP53 ID: 396200	tumor protein p53 [<i>Gallus gallus</i> (chicken)]			
<input type="checkbox"/> TP53	tumor protein	Chromosome E1,		

Recent activity

[Turn Off](#) [Clear](#)

 TP53 AND (alive[prop]) (5574) Gene

 TP53 AND srcdb_refseq[PROP] (6868) Nucleotide

 TP53 (11690) Nucleotide

 Complex Analysis of Retroposed Genes' Contribution to Human

 ORF Capture-Seq as a versatile method for targeted identification

[See more...](#)

TP53 tumor protein p53 [*Homo sapiens* (human)]

Gene ID: 7157, updated on 5-Jul-2020

Summary



Official Symbol	TP53 provided by HGNC
Official Full Name	tumor protein p53 provided by HGNC
Primary source	HGNC:HGNC:11998
See related	Ensembl:ENSG00000141510 MIM:191170
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	P53; BCC7; LFS1; BMFS5; TRP53
Summary	This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. Alternative splicing of this gene and the use of alternate promoters result in multiple transcript variants and isoforms. Additional isoforms have also been shown to result from the use of alternate translation initiation codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016]
Expression	Ubiquitous expression in spleen (RPKM 13.2), lymph node (RPKM 13.1) and 25 other tissues See more
Orthologs	mouse all

Genomic context



Location: 17p13.1

See TP53 in [Genome Data Viewer](#)

Exon count: 12

Table of contents



[Summary](#)

[Genomic context](#)

[Genomic regions, transcripts, and products](#)

[Expression](#)

[Bibliography](#)

[Phenotypes](#)

[Variation](#)

[HIV-1 interactions](#)

[Pathways from PubChem](#)

[Interactions](#)

[General gene information](#)

Markers, Clone Names, Homology,
Gene Ontology

[General protein information](#)

[NCBI Reference Sequences \(RefSeq\)](#)

[Related sequences](#)

[Additional links](#)

Locus-specific Databases

Genome Browsers



[Genome Data Viewer](#)

[Variation Viewer \(GRCh37.p13\)](#)

[Variation Viewer \(GRCh38\)](#)

[1000 Genomes Browser
\(GRCh37.p13\)](#)

Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence: NC_000017.11 Chromosome 17 Reference GRCh38.p13 Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

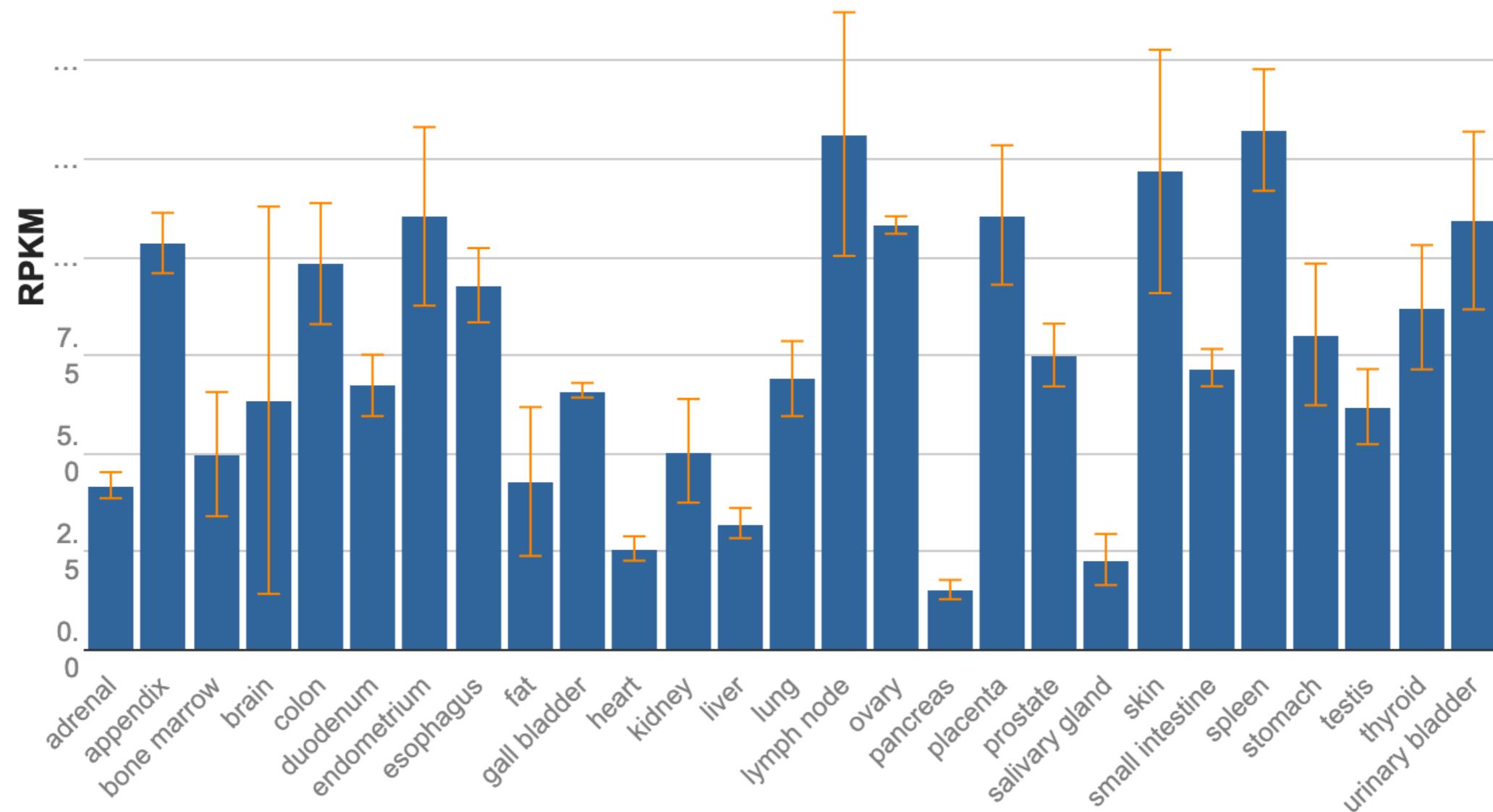


Expression

[See details](#)

HPA RNA-seq normal tissues

- Project title: HPA RNA-seq normal tissues
- Description: RNA-seq was performed of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes
- BioProject: [PRJEB4337](#)
- Publication: [PMID 24309898](#)
- Analysis date: Wed Apr 4 07:08:55 2018



General gene information



Markers

Homology

[Homologs of the TP53 gene](#): The TP53 gene is conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, rat, zebrafish, and frog.

[Orthologs from Annotation Pipeline](#): 241 organisms have orthologs with human gene TP53
[Orthologs](#)

Clone Names

FLJ92943

Gene Ontology [Provided by GOA](#)

Function	Evidence Code	Pubs
ATP binding	IDA	PubMed
DNA binding	IDA	PubMed
DNA binding	IMP	PubMed
DNA-binding transcription activator activity, RNA polymerase II-specific	IDA	PubMed
DNA-binding transcription factor activity	IDA	PubMed
DNA-binding transcription factor activity	IMP	PubMed
DNA-binding transcription factor activity, RNA polymerase II-specific	IDA	PubMed
DNA-binding transcription factor activity, RNA polymerase II-specific	ISA	
DNA-binding transcription factor activity, RNA polymerase II-specific	ISM	PubMed
MDM2/MDM4 family protein binding	IEA	
RNA polymerase II proximal promoter sequence-specific DNA binding	IDA	PubMed
RNA polymerase II proximal promoter sequence-specific DNA binding	ISS	
RNA polymerase II transcription factor binding	IPI	PubMed
TFIID-class transcription factor complex binding	IPI	PubMed



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>.Get the latest research from NIH: <https://www.nih.gov/coronavirus>.Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Display Settings: HomoloGene

Send to:

HomoloGene:460. Gene conserved in Euteleostomi[Download](#) , [Links](#)**Genes***Genes identified as putative homologs of one another during the construction of HomoloGene.*[TP53, *H.sapiens*](#)

tumor protein p53

[TP53, *P.troglodytes*](#)

tumor protein p53

[TP53, *M.mulatta*](#)

tumor protein p53

[TP53, *C.lupus*](#)

tumor protein p53

[TP53, *B.taurus*](#)

tumor protein p53

[Trp53, *M.musculus*](#)

transformation related protein 53

[Tp53, *R.norvegicus*](#)

tumor protein p53

[tp53, *X.tropicalis*](#)

tumor protein p53

[tp53, *D.rerio*](#)

tumor protein p53

Proteins*Proteins used in sequence comparisons and their conserved domain architectures.*[NP_000537.3](#)

393 aa

[XP_001172077.2](#)

393 aa

[NP_001040616.1](#)

393 aa

[NP_001003210.1](#)

381 aa

[NP_776626.1](#)

386 aa

[NP_035770.2](#)

390 aa

[NP_112251.2](#)

391 aa

[NP_001001903.1](#)

362 aa

[NP_001258749.1](#)

374 aa

Protein Alignments*Protein multiple alignment, pairwise similarity scores and***Conserved Domains***Conserved Domains from CDD found in protein sequences*

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. [Explain](#)

Genomic**1. NG_017013.2 RefSeqGene**

Range	5001..24149
Download	GenBank , FASTA , Sequence Viewer (Graphics) , LRG_321

mRNA and Protein(s)**1. NM_000546.6 → NP_000537.3 cellular tumor antigen p53 isoform a**

[See identical proteins and their annotated locations for NP_000537.3](#)

Status: REVIEWED

Description Transcript Variant: This variant (1) can initiate translation from two in-frame AUG start codons. The isoform represented in this variant (a, also known as p53alpha) results from translation initiation at the upstream start codon. Both variants 1 and 2 encode isoform a, which is the longest isoform.

Source sequence(s) [AK223026](#), [DA453049](#), [X02469](#)

Consensus CDS [CCDS11118.1](#)

UniProtKB/Swiss-Prot [P04637](#)

UniProtKB/TrEMBL [K7PPA8](#), [Q53GA5](#)

Related [ENSP00000269305.4](#), [ENST00000269305.8](#)

Conserved Domains (3) [summary](#)

	pfam00870 Location:95 → 289	P53; P53 DNA-binding domain
	pfam07710 Location:319 → 358	P53_tetramer; P53 tetramerisation motif
	pfam08563 Location:5 → 28	P53_TAD; P53 transactivation motif

2. NM_001126112.2 → NP_001119584.1 cellular tumor antigen p53 isoform a



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>.Get the latest research from NIH: <https://www.nih.gov/coronavirus>.Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

GenBank ▾

Send to: ▾

Change region shown

Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

NCBI Reference Sequence: NM_000546.6

[FASTA](#) [Graphics](#)[Go to:](#) ▾

LOCUS	NM_000546	2512 bp	mRNA	linear	PRI	13-MAY-2020
DEFINITION	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA.					
ACCESSION	NM_000546					
VERSION	NM_000546.6					
KEYWORDS	RefSeq; RefSeq Select.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;					
	Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 2512)					
AUTHORS	Marcel V, Tran PL, Sagne C, Martel-Planche G, Vaslin L,					
	Teulade-Fichou MP, Hall J, Mergny JL, Hainaut P and Van Dyck E.					
TITLE	G-quadruplex structures in TP53 intron 3: role in alternative					
	splicing and in production of p53 mRNA isoforms					
JOURNAL	Carcinogenesis 32 (3), 271-278 (2011)					
PUBMED	21112961					
REFERENCE	2 (bases 1 to 2512)					
AUTHORS	Marcel V, Perrier S, Aoubala M, Ageorges S, Groves MJ, Diot A,					

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Articles about the TP53 gene

[Cerebrospinal fluid TP53 gene expression in patients [Zhonghua Yi Xue Za Zhi 2011;90(10):631-635]

SIRT3 increases cisplatin sensitivity in small-cell lung cancer through [Genes 2011;2(3):311-321]

Evaluation of p53 and Its Target Expressions [Asian Pac J Cancer Prev 2011;12(10):2751-2755]

Nucleotide

Nucleotide

Advanced

Search



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>.

Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

FASTA ▾

Send to: ▾

Change region shown

Customize view

Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

NCBI Reference Sequence: NM_000546.6

[GenBank](#) [Graphics](#)

>NM_000546.6 Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCCTCGGGC
TGGGAGCGTGCTTCCACGACGGTGACACGCTCCCTGGATTGGCAGCCAGACTGCCTCCGGGTCACTG
CCATGGAGGAGCCGCAGTCAGATCCTAGCGTCAGCCCCCTCTGAGTCAGGAAACATTTCAGACCTATG
GAAACTACTTCCCTGAAAACAACGTTCTGTCCCCCTGCCGTCCAAGCAATGGATGATTGATGCTGTCC
CCGGACGATATTGAACAATGGTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTG
CTCCCCCGTGGCCCTGCACCAGCAGCTCCTACACCGCGGCCCTGCACCAGCCCCCTGGCCCT
GTCATCTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTACGGTTCCGTCTGGGCTTGCATTCT
GGGACAGCCAAGTCTGTGACTTGCACGTACTCCCTGCCCTAACAAAGATGTTGCCAACTGCCAAGA
CCTGCCCTGTGCAGCTGGTTGATTCCACACCCCCGCCGGACCCCGTCCGCCATGCCATCTA
CAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCCTGCCCTCACCATGAGCGCTGCTCAGATAGCGAT
GGTCTGGCCCTCCTCAGCATTTATCCGAGTGGAAAGGAAATTGCGTGTGGAGTATTGGATGACAGAA
ACACTTTGACATAGTGTGGTGGTGCCTATGAGCCGCTGAGGTTGGCTCTGACTGTACCACCATCCA
CTACAACATGTGTAACAGTTCCTGCATGGCGGCATGAACCGGAGGCCATCCTACCATCATCACA
CTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTGAGGTGCGTGTGCTGTCCCTGGGA
GAGACCGCGCACAGAGGAAGAGAAATCTCGCAAGAAAGGGAGCCTCACACAGCTGCCCTCAGGGAG
CACTAAGCGAGCACTGCCAACACACCAGCTCCTCTCCCAGCCAAAGAAGAAACCAGTGGATGGAGAA
TATTCACCTTCAGATCCGTGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCCTGGAAC
TCAAGGATGCCAGGCTGGGAAGGAGCCAGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAA
GGGTCACTACCTCCGCCATAAAACTCATGTTCAAGACAGAAGGGCCTGACTCAGACTGACATTCT
CCACTTCTGTTCCCCACTGACAGCCTCCCACCCCATCTCTCCCTCCCTGCCATTGGGTTGGGT
CTTTCAACCCCTTCCATTACCTCTCCCTCAACACCCACCCACTTCCATTTCCTTCTCCCCCCCC

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Articles about the TP53 gene

[Cerebrospinal fluid TP53 gene mu
in patien [Zhonghua Yi Xue Za Zhi.

SIRT3 increases cisplatin sensitivit
small-cell lung cancer thro [Gene.

Evaluation of p53 and Its Target Ge
Express [Asian Pac J Cancer Prev.

Se



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>.

Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Standard Nucleotide BLAST

blastn **blastp** **blastx** **tblastn** **tblastx**

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

[Reset page](#)

[Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

NM_000546.6

[Clear](#)

Query subrange

From

To

Or, upload file

[Browse...](#)

No file selected.

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

BLAST results will be displayed

in a new format by default

You can always switch back to the

Traditional Results page.



Choose Search Set

Database

Standard databases (nr etc.): rRNA/ITS databases Genomic + transcript databases Betacoronavirus

Nucleotide collection (nr/nt)

SRA (sequence read archive)

www.ncbi.nlm.nih.gov/sra

NCBI Resources How To

swheelan My NCBI Sign Out

SRA SRA Advanced Search Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.



SRA - Now available on the cloud

Sequence Read Archive (SRA) data, available through multiple cloud providers and NCBI servers, is the largest publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.

Announcement

[NIH Request for Information \(RFI\) on SRA data format changes and plans.](#)

Getting Started

[How to Submit](#)

[How to search and download](#)

[How to use SRA in the cloud](#)

[Submit to SRA](#)

Tools and Software

[Download SRA Toolkit](#)

[SRA Toolkit Documentation](#)

[SRA-BLAST](#)

[SRA Run Browser](#)

[SRA Run Selector](#)

Related Resources

[Submission Portal](#)

[Trace Archive](#)

[dbGaP Home](#)

[BioProject](#)

[BioSample](#)

SRA

SRA

human neuron

Search

Create alert Advanced

Help



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>.Get the latest research from NIH: <https://www.nih.gov/coronavirus>.Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Access

Controlled (8,713)

Public (14,182)

Source

DNA (3,478)

RNA (18,920)

Type

genome (1,797)

Library Layout

paired (3,422)

single (19,474)

Platform

Helicos (51)

Illumina (22,797)

Ion Torrent (31)

PacBio SMRT (17)

Strategy

EpiGenomics (288)

Exome (97)

Genome (2,384)

RNASeq (6)

other (20,121)

Data in Cloud

GS (22,861)

S3 (22,867)

Summary ▾ 20 per page ▾

Send to: ▾

Filters: [Manage Filters](#)

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Search results

Items: 1 to 20 of 22896

<< First < Prev Page 1 of 1145 Next > Last >>

 [GSM4632031: 3'-tag RNA-seq PSAPKO neuron; Homo sapiens; RNA-Seq](#)

1. 3 ILLUMINA (Illumina HiSeq 4000) runs: 13.5M spots, 1.4G bases, 585Mb downloads
Accession: SRX8596426

 [GSM4632030: 3'-tag RNA-seq WT neuron; Homo sapiens; RNA-Seq](#)

2. 3 ILLUMINA (Illumina HiSeq 4000) runs: 12.4M spots, 1.2G bases, 539.4Mb downloads
Accession: SRX8596425

 [GSM4632028: CROP-seq_CRISPRa_10x_lane_2; Homo sapiens; RNA-Seq](#)

3. [CROP-seq_CRISPRa_10x_lane_2](#)
1 ILLUMINA (Illumina NovaSeq 6000) run: 663.2M spots, 84.2G bases, 29.9Gb downloads
Accession: SRX8596424

 [GSM4632027: CROP-seq_CRISPRa_10x_lane_1; Homo sapiens; RNA-Seq](#)

4. [CROP-seq_CRISPRa_10x_lane_1](#)
1 ILLUMINA (Illumina NovaSeq 6000) run: 693.5M spots, 88.1G bases, 31.6Gb downloads
Accession: SRX8596423

Results by taxon

Top Organisms [Tree]

Homo sapiens (20467)

Mus musculus (2363)

mixed sample (41)

Rattus norvegicus (18)

Pan troglodytes (7)

Top Bioprojects

Production ENCODE functional... (27)

Production ENCODE epigenomic... (2)

Mouse ENCODE functional geno... (2)

Search in related databases

Database	Access		all
	public	controlled	
BioSample	6,757	8,268	15,025
BioProject	222	7	229
dbGaP			11
GEO Datasets	17,724		17,724

SRA

- many organisms included, but largely human
- all sequencing modalities and technologies represented
- fairly consistent annotation
- not all records are publicly available

SRA-BLAST

Sequence Read Archive Nucleotide BLAST

blastn

BLASTN programs search SRA databases using a nucleotide query. [?](#)

[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

[Clear](#) **Query subrange** [?](#)

From To

Or, upload file [Browse...](#) No file selected. [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

BLAST results will be displayed in a new format by default

You can always switch back to the Traditional Results page.



Choose Search Set

SRA Experiment set (SRX)

Enter an SRA accession (experiment, study, or submission), title, the scientific name or tax id—completion: [+](#)
Enter an SRA accession (experiment, study, or submission), title, the scientific name or tax id. Only 20 top suggestions will be shown.
[?](#)

Program Selection

Optimize for

Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

BLAST Search database **SRA** using **Megablast (Optimize for highly similar sequences)**

other NCBI resources

- PubMed
- tools (Cn3D, Splign, COBALT, OSIRIS, Genome Workbench etc)
- hundreds of specialized databases and resources (food safety, MHC, HIV, influenza, OMIM etc)
- tutorials <https://www.ncbi.nlm.nih.gov/home/learn/>

Ensembl and BioMart

- Ensembl is a database and genome browser that focuses on vertebrate genomes
- browser has extensive tools for exploring variation

Newest updates include:

- filtering and prioritization of genome variation
- Manhattan plot visualization for linkage disequilibrium and eQTL data
- ontology search for phenotypes, traits and disease

Tools**BioMart >**[All tools](#)

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 100 (April 2020)

- Update to GENCODE 34 (human) and GENCODE M25 (mouse)
- Update of gnomAD genomic allele frequencies to version 3
- New genomes: 3 mammals, 7 fish, 6 birds, 4 reptiles
- Updated genomes: Platypus and Northern Pike
- New interface for configuration of multidimensional track hubs

Search

All species



for



e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

All genomes -- Select a species --**Pig breeds**

Pig reference genome and 12 additional breeds

[View full list of all species](#)**Favourite genomes** **Human**

GRCh38.p13

**Mouse**

GRCm38.p6

**Zebrafish**

GRCz11

[More release news](#)  on our blog**Other news from our blog**

- 26 Jun 2020: [Cool stuff the Ensembl VEP can do: analysis with RefSeq transcripts](#) 
- 25 Jun 2020: [Ensembl Rapid Release](#) 
- 12 Jun 2020: [What's coming up in Ensembl 101 / Ensembl Genomes 48?](#) 

New Search

Current selection:

< all Species

Only searching Human

Restrict category to:

Gene	468
Transcript	1245
Variant	299
Phenotype	1
GeneTree	17
Clones & Regions	1
Protein Domain	1
Protein Family	11

Per page:

10 25 50 100

Layout:

Standard  Table

Tip:

You can choose which results appear near the top of your search by updating your favourite species.

Only searching Human ▾ tp53

2043 results match tp53 when restricted to species: Human [TP53 \(Human Gene\)](#)**ENSG00000141510** 17:7661779-7687550:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]

LRG_321 (LRG display in Ensembl gene record; description: Locus Reference Genomic record for **TP53**,) is an external reference matched to Gene ENSG00000141510[Variant table](#) • [Phenotypes](#) • [Location](#) • [External Refs.](#) • [Regulation](#) • [Orthologues](#) • [Gene tree](#)[TP53-222 \(Human Transcript\)](#)**ENST00000615910** 17:7669609-7676594:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].

[Location](#) • [External Refs.](#) • [cDNA seq.](#) • [Exons](#) • [Variant table](#) • [Protein seq.](#) • [Population](#) • [Protein summary](#)[TP53-223 \(Human Transcript\)](#)**ENST00000617185** 17:7668402-7687550:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]

LRG_321t3 (LRG display in Ensembl transcript record; description: Locus Reference Genomic record for **TP53**) is an external reference matched to Transcript ENST00000617185[Location](#) • [External Refs.](#) • [cDNA seq.](#) • [Exons](#) • [Variant table](#) • [Protein seq.](#) • [Population](#) • [Protein summary](#)[TP53-215 \(Human Transcript\)](#)**ENST00000571370** 17:7685260-7686371:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].

[Location](#) • [External Refs.](#) • [cDNA seq.](#) • [Exons](#) • [Variant table](#) • [Population](#)[TP53-216 \(Human Transcript\)](#)**ENST00000574684** 17:7674254-7675119:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].

[Location](#) • [External Refs.](#) • [cDNA seq.](#) • [Exons](#) • [Variant table](#) • [Population](#)[TP53-217 \(Human Transcript\)](#)**ENST00000576024** 17:7669569-7673587:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].

[Location](#) • [External Refs.](#) • [cDNA seq.](#) • [Exons](#) • [Variant table](#) • [Protein seq.](#) • [Population](#) • [Protein summary](#)[TP53-229 \(Human Transcript\)](#)**ENST00000635293** 17:7665416-7687491:-1

Tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998].

[Location](#) • [External Refs.](#) • [cDNA seq.](#) • [Exons](#) • [Variant table](#) • [Protein seq.](#) • [Population](#) • [Protein summary](#)

In this section

- Control Panel
- Find a Data Display
- Adding Custom Tracks
 - BED File Format
 - Large File Formats
 - Pairwise Interactions File Format
 - GFF/GTF File Format
 - GFF3 File Format
 - PSL File Format
 - Variation File Format
 - WIG File Format
 - Coordinate Systems for custom tracks
- Adding Track Hubs
- Public Track Hubs
- Track Hub Support in Ensembl
- Tutorials
 - Retrieving sequences
 - Gene Expression
 - Compare genes across species
 - Variants for my gene
 - Diseases and Phenotypes
 - ENCODE data in Ensembl
 - The GRCh37 assembly in Ensembl
 - Use my own data
- Glossary
- Supported browsers
- Browse News by Topic
- Archives
 - Archives: Table of assemblies

Search documentation...

Go

[Home](#) > [Help & Documentation](#) > [Using this website](#) > [Find a Data Display](#)

Find a Data Display

Not sure how to find the data visualisation you need? Choose the type of data you are interested in, then browse a selection of relevant visualisations!

Species:Human **Feature type:**

- Genes
- Genomic locations
- Variants

Identifier:

ENSG00000139618

Go

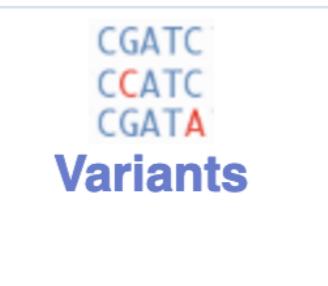


Ensembl release 100 - April 2020 © EMBL-EBI

[Permanent link](#)

About Us[About us](#)[Contact us](#)[Citing Ensembl](#)[Privacy policy](#)[Disclaimer](#)**Get help**[Using this website](#)[Adding custom tracks](#)[Downloading data](#)[Video tutorials](#)[Variant Effect Predictor \(VEP\)](#)**Our sister sites**[Ensembl Bacteria](#)[Ensembl Fungi](#)[Ensembl Plants](#)[Ensembl Protists](#)[Ensembl Metazoa](#)**Follow us** [Blog](#) [Twitter](#) [Facebook](#)

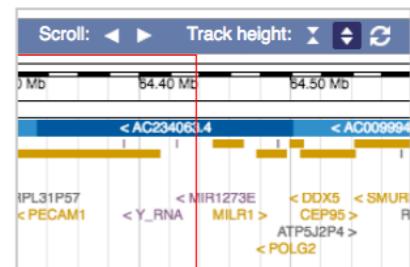
47 views for Gene data

**Sequence & Structure****Expression & Regulation****Transcripts & Proteins****Comparative Genomics****Variants****Sequence & Structure displays for:**

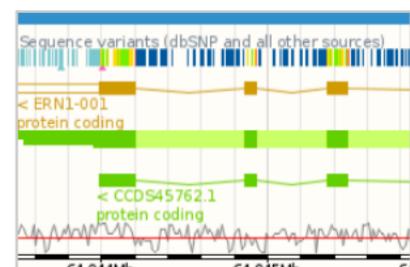
ENSG00000139618

Update**Scrolling Browser**

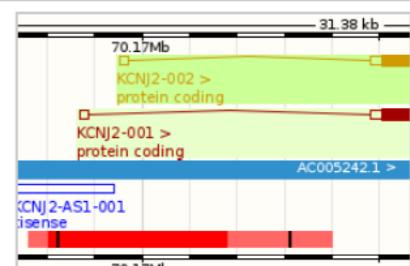
View the position of this gene in our fully scrollable genome browser

Show me**Region in Detail**

Zoom in on your gene of interest

Show me**Immediate Neighbourhood**

View this gene in its genomic location

Show me

BioMart

- enables complex database queries (e.g. “all genes in XXX region with an annotated variant, that are also in YYY pathway”)
- need to specify database, filter (how to restrict the search), and attributes (what you want returned)
- web-based, Perl, and R interfaces are available

[!\[\]\(ec1b4bedfa6077be5e53bf0276b8c0c5_img.jpg\) New](#) [!\[\]\(ad073283f15b3fe848e7fc450f454498_img.jpg\) Count](#) [!\[\]\(1fdbdcb701bacdea294156907d5ff57f_img.jpg\) Results](#)[!\[\]\(7487782dbad083afbf914fa6b7731eae_img.jpg\) URL](#) [!\[\]\(1b4185381a2d93e8f2a2e8edd581382c_img.jpg\) XML](#) [!\[\]\(537f91af50728829ee70d35dd2d9b2c3_img.jpg\) Perl](#) [!\[\]\(dcbedb6639942853bf6a1d35e7255d7a_img.jpg\) Help](#)**Dataset**

[None selected]

- CHOOSE DATABASE - 

simple example: fetch all variants for human genes on chromosome 19

 east

BLAST/BLAT | VEP | Tools | BioMart | Downloads | More ▾  Search all species... 

[Login/Register](#)

[New](#) [Count](#) [Results](#) [URL](#) [XML](#) [Perl](#) [Help](#)

Dataset
Human genes (GRCh38.p13)

Filters
[None selected]

Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

+ REGION:
+ GENE:
+ PHENOTYPE:
+ GENE ONTOLOGY:
+ MULTI SPECIES COMPARISONS:
+ PROTEIN DOMAINS AND FAMILIES:
+ VARIANT:

[New](#) | [Count](#) | [Results](#)[URL](#) | [XML](#) | [Perl](#) | [Help](#)**Dataset**

Human genes (GRCh38.p13)

Filters

Chromosome/scaffold: 19

Attributes

Gene stable ID

Gene stable ID version

Transcript stable ID

Transcript stable ID version

Dataset

[None Selected]

Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION: Chromosome/scaffold

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19**
- 20

 Coordinates

Start

End

 Multiple regions (Chr:Start:End:Strand) [Max 500 advised]

e.g. 1:100:10000:-1, 1:100000:200000:1

Browse... No file selected.**GENE:**

 New Count Results URL XML Perl Help**Dataset**

Human genes (GRCh38.p13)

Filters

Chromosome/scaffold: 19

Attributes

Gene stable ID

Gene stable ID version

Variant name

Dataset

[None Selected]

Please select columns to be included in the output and hit 'Results' when ready**Missing non coding genes in your mart query output, please check the following [FAQ](#)** Features Structures Homologues (Max select 6 orthologues) Variant (Germline) Variant (Somatic) Sequences**GENE:****Ensembl** Gene stable ID Gene stable ID version Version (gene) Transcript stable ID Transcript stable ID version Version (transcript) Protein stable ID Protein stable ID version Version (protein) Chromosome/scaffold name Gene start (bp) Gene end (bp) Strand Karyotype band Gene name Source of gene name Transcript count Gene % GC content Gene description**GERMLINE VARIANT INFORMATION:****Variant** Variant name Variant source Variant source description Variant alleles Variant supporting evidence Mapweight Minor allele Minor allele frequency Minor allele count Clinical significance**Variants Location** Transcript location (bp) Variant chromosome Strand Protein location (aa) chromosome/scaffold position start (bp) Chromosome/scaffold position end (bp)

[New](#) [Count](#) [Results](#)[★ URL](#) [XML](#) [Perl](#) [Help](#)**Dataset**

Human genes (GRCh38.p13)

Filters

Chromosome/scaffold: 19

Attributes

Gene stable ID

Gene stable ID version

Variant name

Dataset

[None Selected]

Export all results to

File

only

 Go

TSV

Unique results

Email notification to

View

10

rows as

HTML

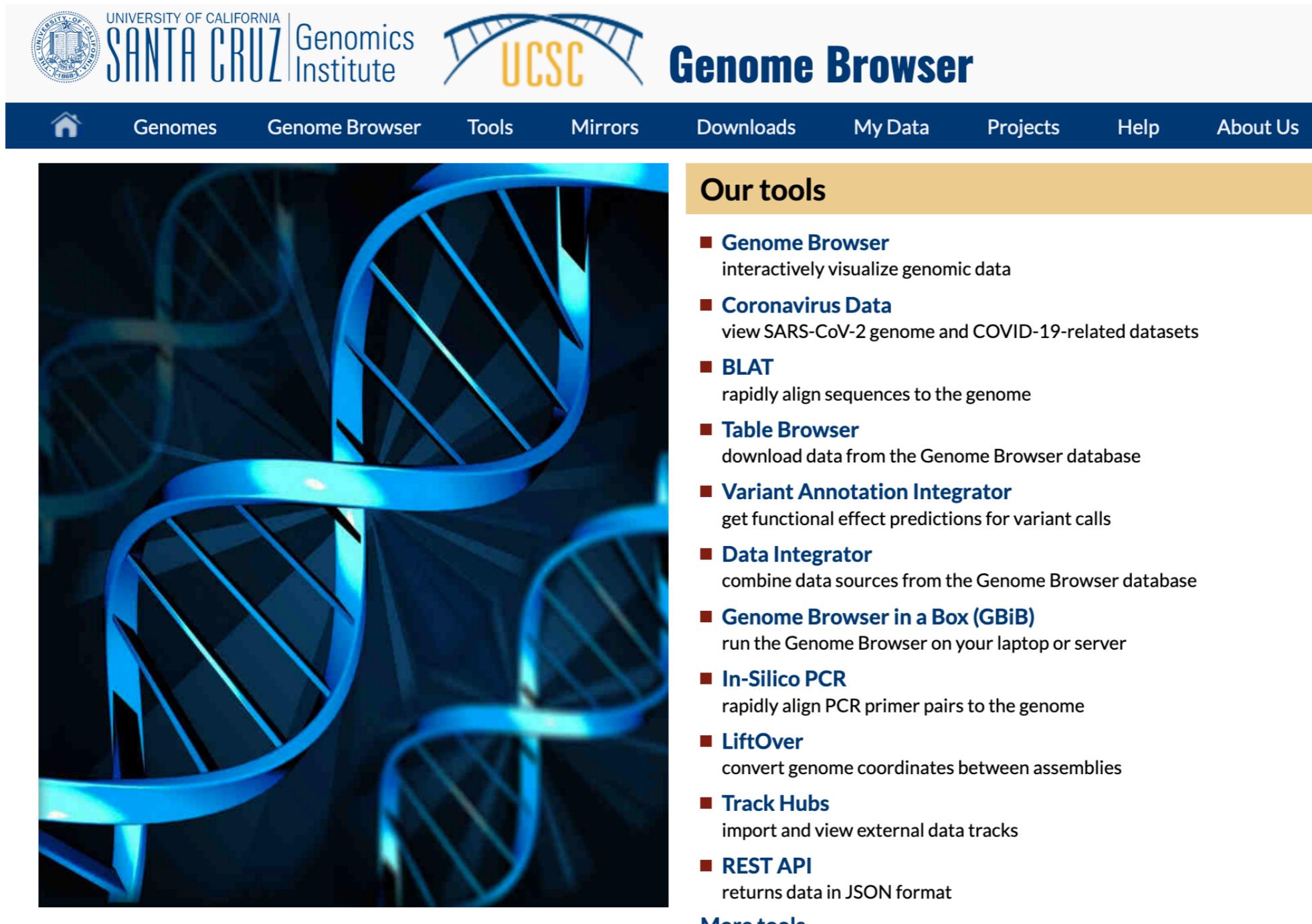
Unique results only

Gene stable ID	Gene stable ID version	Variant name
ENSG00000264910	ENSG00000264910.2	rs961386574
ENSG00000264910	ENSG00000264910.2	rs1031830492
ENSG00000264910	ENSG00000264910.2	rs1031830492
ENSG00000264910	ENSG00000264910.2	rs111287428
ENSG00000264910	ENSG00000264910.2	rs111287428
ENSG00000264910	ENSG00000264910.2	rs956139889
ENSG00000264910	ENSG00000264910.2	rs193231731
ENSG00000264910	ENSG00000264910.2	rs868499727
ENSG00000264910	ENSG00000264910.2	rs1378305240
ENSG00000264910	ENSG00000264910.2	rs1455718510

UCSC Genome Browser and Table Browser

genome.ucsc.edu

- hosts a treasure trove of data and annotation



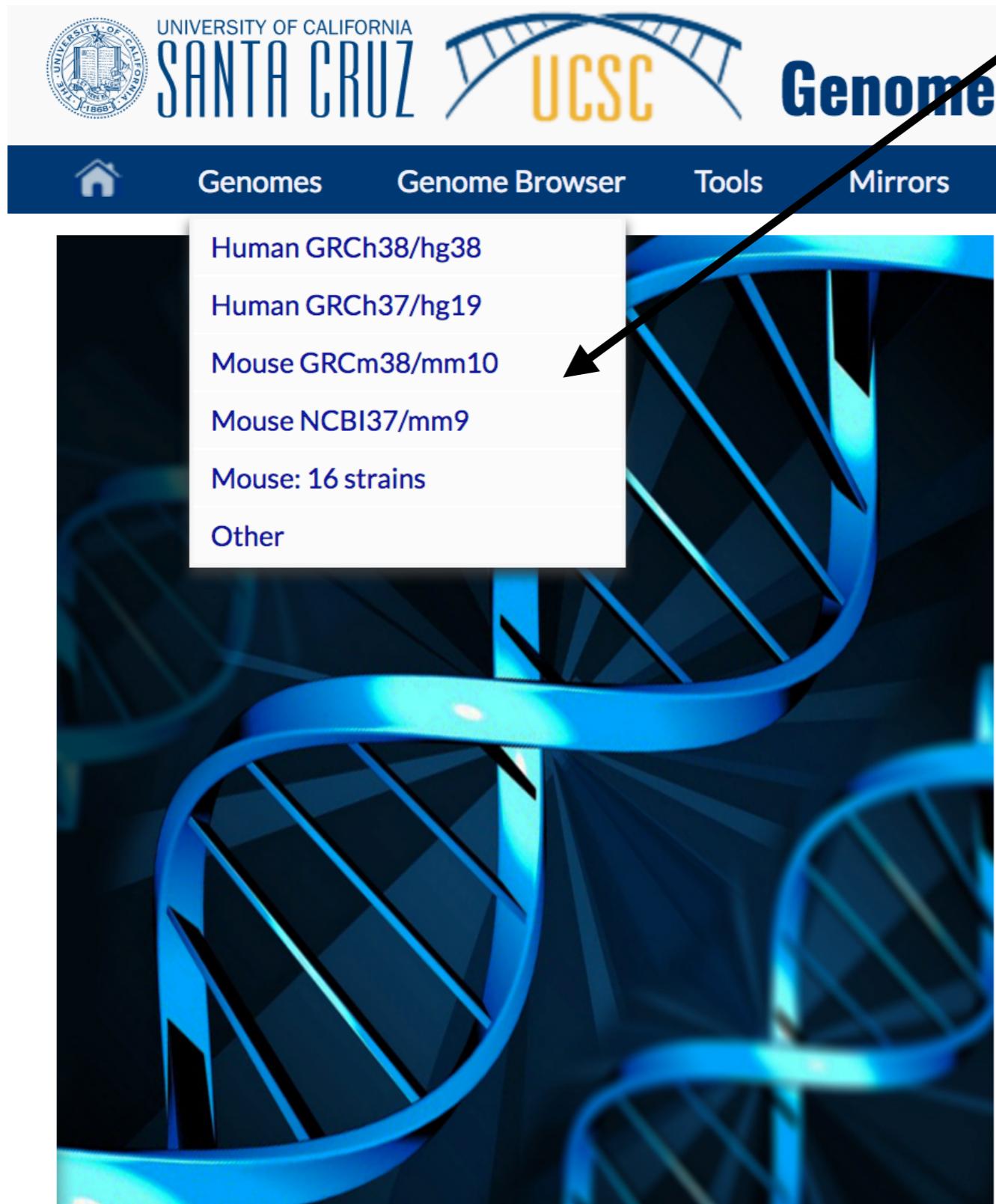
The screenshot shows the UCSC Genome Browser homepage. At the top left is the UCSC Genomics Institute logo. To its right is the UCSC logo. The main title "Genome Browser" is centered above a navigation bar. The navigation bar includes links for Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. Below the navigation bar is a large blue banner featuring a 3D rendering of several glowing blue DNA double helix molecules against a dark background. To the right of the banner is a yellow sidebar titled "Our tools" which lists various genomic analysis tools.

Our tools

- **Genome Browser**
interactively visualize genomic data
- **Coronavirus Data**
view SARS-CoV-2 genome and COVID-19-related datasets
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **Track Hubs**
import and view external data tracks
- **REST API**
returns data in JSON format

[More tools...](#)

<http://genome.ucsc.edu/>



be careful about genome build! An assembly (or “build”) is a fixed version of a genome. Builds are released every several years.

You should always be aware whether you are using GRCh38 (hg38) or GRCh37 (hg19) for the human genome. They have different coordinates and are annotated with different types of information such as experimental data sets.

UCSC Genome Browser

- lots of tracks available for each genome and build! function, structure, conservation, additional datasets
- you can upload custom tracks
- you can hide/show and change display configuration for tracks

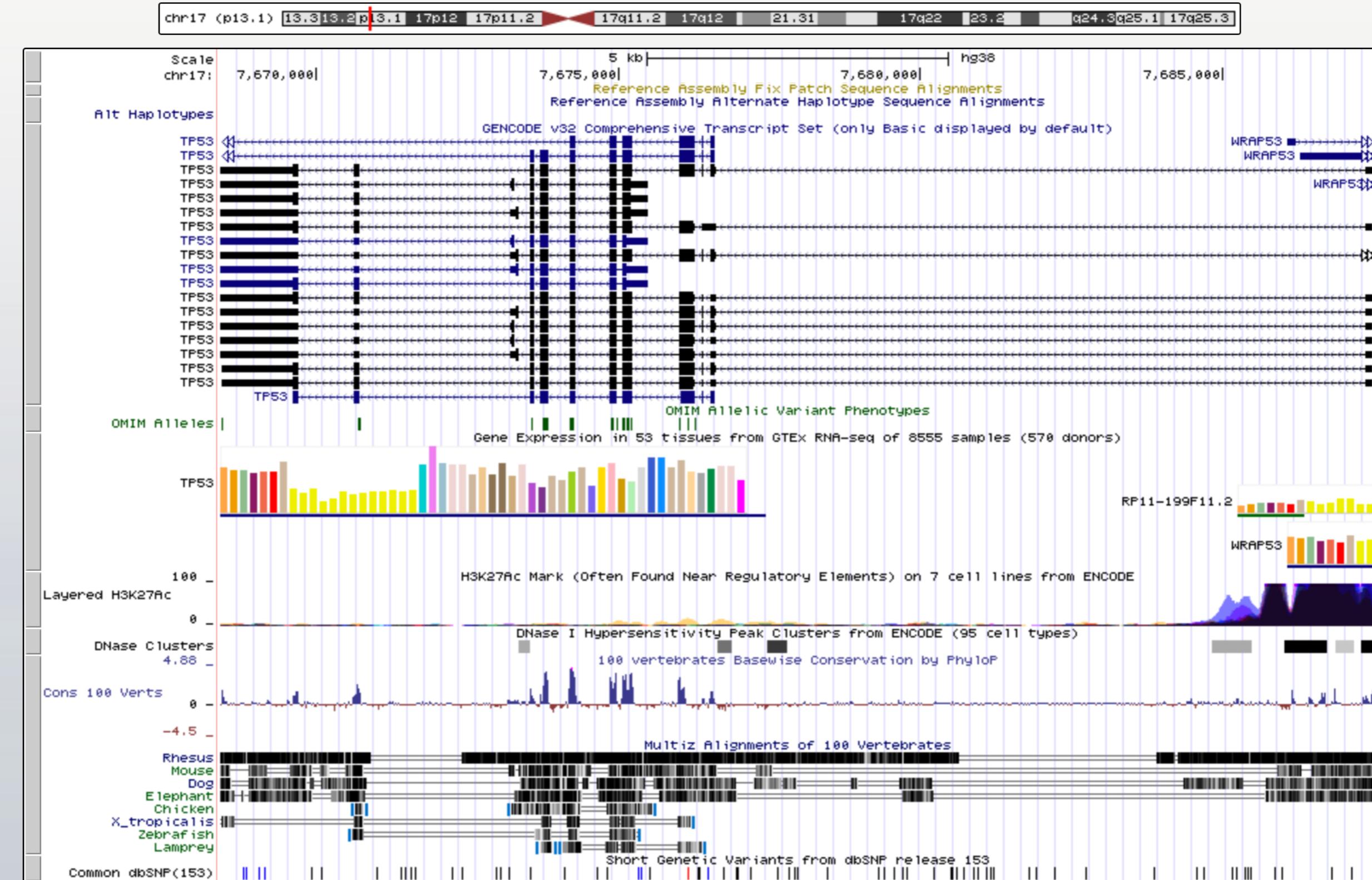
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr17:7,668,402-7,687,538 19,137 bp.

enter position, gene symbol, HGVS or search terms

go



here you can access all of
the data used to make the
Genome Browser tracks!

The image shows the UCSC Genome Browser homepage. A black arrow points from the text above to the "Tools" menu item in the top navigation bar. The "Tools" menu is currently open, displaying a list of tools:

- Blat
- Table Browser
- Variant Annotation Integrator
- Data Integrator
- Gene Interactions
- Gene Sorter
- Genome Graphs
- In-Silico PCR
- LiftOver
- VisiGene
- Other Utilities

The main content area features a large blue DNA helix graphic. To the right of the tools menu, there is a section titled "tools" which lists several additional tools:

- Genome Browser**: interactively visualize genomic data
- Coronavirus Data**: SARS-CoV-2 genome and COVID-19-related datasets
- BLAT**: rapidly align sequences to the genome
- Table Browser**: download data from the Genome Browser database
- Variant Annotation Integrator**: functional effect predictions for variant calls
- Data Integrator**: combine data sources from the Genome Browser database
- Genome Browser in a Box (GBiB)**: run the Genome Browser on your laptop or server
- In-Silico PCR**: rapidly align PCR primer pairs to the genome
- LiftOver**: convert genome coordinates between assemblies
- Track Hubs**: import and view external data tracks
- REST API**: returns data in JSON format

[More tools...](#)

Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed

What's new

Jul. 07, 2020 - [New SARS-CoV-2 data; Human genome turns 20](#)

what annotated repeats are near the TP53 gene?

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

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. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you may want to use [Galaxy](#) or our [public API](#). If you have a set of tracks that you would like to analyze through annotation enrichments, send the data to [GREAT](#). Refer to the [FAQ](#) for more information about what data is associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38)

group: Repeats track: RepeatMasker add custom tracks track hubs

table: rmsk **describe table schema**

region: genome position chr17:7,668,402-7,687,538 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

output format: all fields from selected table Send output to Galaxy GREAT

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

get output summary/statistics

To reset **all** user cart settings (including custom tracks), [click here](#).

repeatmasker annotated repeats

type the gene name, click "lookup"

"describe table schema" will tell you what data are available in the selected tables



Schema for RepeatMasker - Repeating Elements by RepeatMasker

Database: hg38 **Primary Table:** rmsk **Row Count:** 5,607,738 **Data last updated:** 2018-08-10

Format description: RepeatMasker .out record

field	example	SQL type	description
bin	585	smallint(5) unsigned	Indexing field to speed chromosome range queries.
swScore	463	int(10) unsigned	Smith Waterman alignment score
milliDiv	13	int(10) unsigned	Base mismatches in parts per thousand
milliDel	6	int(10) unsigned	Bases deleted in parts per thousand
milliIns	17	int(10) unsigned	Bases inserted in parts per thousand
genoName	chr1	varchar(255)	Genomic sequence name
genoStart	10000	int(10) unsigned	Start in genomic sequence
genoEnd	10468	int(10) unsigned	End in genomic sequence
genoLeft	-248945954	int(11)	-#bases after match in genomic sequence
strand	+	char(1)	Relative orientation + or -
repName	(TAACCC)n	varchar(255)	Name of repeat
repClass	Simple_repeat	varchar(255)	Class of repeat
repFamily	Simple_repeat	varchar(255)	Family of repeat
repStart	1	int(11)	Start (if strand is +) or -#bases after match (if strand is -) in repeat sequence
repEnd	471	int(11)	End in repeat sequence
repLeft	0	int(11)	-#bases after match (if strand is +) or start (if strand is -) in repeat sequence
id	1	char(1)	First digit of id field in RepeatMasker .out file. Best ignored.

Sample Rows

bin	swScore	milliDiv	milliDel	milliIns	genoName	genoStart	genoEnd	genoLeft	strand	repName	repClass	repFan
585	463	13	6	17	chr1	10000	10468	-248945954	+	(TAACCC)n	Simple_repeat	Simple_r
585	3612	114	215	13	chr1	10468	11447	-248944975	-	TAR1	Satellite	telo
585	484	251	132	0	chr1	11504	11675	-248944747	-	L1MC5a	LINE	L1
585	239	294	19	10	chr1	11677	11780	-248944642	-	MER5B	DNA	hAT-Chai
585	219	230	27	0	chr1	15261	15255	248941067		MID2	SINE	MID

what annotated repeats are near the TP53 gene?

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

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. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you may want to use [Galaxy](#) or our [public API](#). If you have a set of tracks that you would like to analyze further through annotation enrichments, send the data to [GREAT](#). Refer to the [FAQ](#) for more information.

RepeatMasker annotated repeats

entirely from the [Sequence and Annotation Downloads page](#).

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38)
group: Repeats track: RepeatMasker add custom tracks track hubs
table: rmsk describe table schema
region: genome position chr17:7,668,402-7,687,538 lookup define regions
identifiers (names/accessions): paste list upload list
filter: create
intersection: create
output format: all fields from selected table Send output to Galaxy GREAT
output file: (leave blank to keep output in browser)
file type returned: plain text gzip compressed

get output summary/statistics

To reset **all** user cart settings (including custom tracks), [click here](#).

type the gene name, click “lookup”

A red circle highlights the "get output" button at the bottom left of the form. A red arrow points from the text "To reset **all** user cart settings (including custom tracks), [click here](#)." up towards the "get output" button.

#bin	swScore	milliDiv	milliDel	milliIns	genoName	genoStart	genoEnd	genoLeft	strand
repName	repClass	repFamily	repStart	repEnd	repLeft	id			
643	2031	159	10	0	chr17	7668548	7668856	-75588585	+
0	1								AluJb
643	299	296	26	11	chr17	7669141	7669244	-75588197	-
110	1								MIRb
643	2264	80	23	13	chr17	7669895	7670197	-75587244	-
1	1								AluSp
643	203	292	50	22	chr17	7670203	7670293	-75587148	-
49	1								MIRc
643	521	221	0	19	chr17	7670752	7670858	-75586583	+
107	-259	1							MER47A
643	2140	138	0	0	chr17	7670873	7671171	-75586270	-
1	1								AluSx
643	610	221	3	23	chr17	7671172	7671300	-75586141	-
6	1								FLAM_C
643	2406	77	7	0	chr17	7671300	7671599	-75585842	-
1	1								AluSx1
643	427	229	113	0	chr17	7671599	7671635	-75585806	-
87	1								AluJb
643	923	119	0	26	chr17	7671660	7671815	-75585626	+
2290	-116	1							Tigger5
643	1690	192	30	3	chr17	7671815	7672107	-75585334	+
-12	1								DNA
643	2312	71	7	0	chr17	7672110	7672392	-75585049	+
-27	1								AluJo
643	27	0	53	25	chr17	7672392	7672430	-75585011	+
Simple_repeat	1	40	0	1					(AAAAT)n
643	965	162	0	25	chr17	7672430	7672590	-75584851	+
-24	1								Simple_repeat
643	1314	165	27	14	chr17	7672598	7672816	-75584625	+
-11	1								AluJb
643	2300	79	0	48	chr17	7672828	7673144	-75584297	+
-10	1								AluSx
643	844	112	0	0	chr17	7673144	7673269	-75584172	+
348	472	0	1						Tigger5b
643	2395	60	7	13	chr17	7674360	7674659	-75582782	-
1	1								AluY
643	307	316	30	0	chr17	7674668	7674766	-75582675	+
-66	1								MIRc
643	2117	146	0	13	chr17	7675388	7675697	-75581744	-
1	1								AluJb
643	12	81	36	35	chr17	7676466	7676494	-75580947	+
1	29	0	1						G-rich
643	210	298	123	45	chr17	7676758	7676937	-75580504	+
-1810	1								L2
643	2217	118	3	0	chr17	7676938	7677234	-75580207	-
2	1								AluSx
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								L2a
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								L2
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3219
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-75579637	+
-17	1								3409
643	2217	118	3	0	chr17	7677629	7677804	-75579637	+
2	1								3409
643	337	238	120	40	chr17	7677629	7677804	-755796	

what SNPs are in the TP53 gene?

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

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. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)

group: Variation **track:** dbSNP 153 [add custom tracks](#) [track hubs](#)

table: Common dbSNP(153) (dbSnp153Common) [describe table schema](#)

region: genome position chr17:7,668,402-7,687,538 [lookup](#) [define regions](#)

identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

subtrack merge: [create](#)

intersection: [create](#)

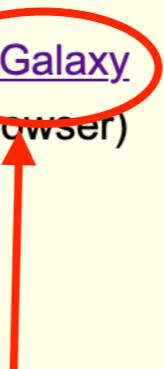
output format: BED - browser extensible data [Send output to](#) [Galaxy](#) [GREAT](#)

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

[get output](#) [summary/statistics](#)

To reset **all** user cart settings (including custom tracks), [click here](#).



[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

Output SNP150Common as BED

Include custom track header:

name= tb_snp150Common

description= table browser query on SNP150Common

visibility= pack ▾

url=

Create one BED record per:

Whole Gene

Upstream by 1 bases

Downstream by 200 bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

[Send query to Galaxy](#)

[Cancel](#)

BED format is a simple way to store feature locations

3 required columns:

chrom

chromStart

chromEnd

coordinates are zero-based.



Frequently Asked Questions: Data File Formats

Topics

General formats

- [Axt format](#)
- [BAM format](#)
- [BED format](#)
- [BED detail format](#)
- [bedGraph format](#)
- [barChart and bigBarChart format](#)
- [bigBed format](#)
- [bigGenePred table format](#)
- [bigPsl table format](#)
- [bigMaf table format](#)
- [bigChain table format](#)
- [bigNarrowPeak table format](#)
- [bigWig format](#)
- [Chain format](#)
- [CRAM format](#)
- [GenePred table format](#)
- [GFF format](#)
- [GTF format](#)
- [HAL format](#)
- [Hic format](#)
- [Interact and bigInteract format](#)
- [MAF format](#)
- [Microarray format](#)
- [Net format](#)
- [Personal Genome SNP format](#)
- [PSL format](#)
- [VCF format](#)
- [WIG format](#)

ENCODE-specific formats

- [ENCODE broadPeak format](#)
- [ENCODE gappedPeak format](#)
- [ENCODE narrowPeak format](#)
- [ENCODE pairedTagAlign format](#)
- [ENCODE peptideMapping format](#)

Galaxy

- accessible to users with or without a programming background
- focus on NGS, sequence analysis tools
- main site is <http://usegalaxy.org>, available to anyone. Storage and computational quotas apply and are higher if using a registered account
- many other Galaxy servers exist! more than 80 publicly accessible sites, typically specialized for a biological problem (RNAseq, metagenomics, whale sharks, insects, ChIP etc)
- you can download the software and set up a local Galaxy instance, if you want to use your own tools and hardware

tool list

workspace

history

Galaxy Analyze Data Workflow Visualize ▾ Shared Data ▾ Help ▾ Login or Register Using 0%

All tools should be functioning normally with the exception of RNA STAR.

Tools

search tools

Get Data
Collection Operations
GENERAL TEXT TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Datamash
GENOMIC FILE MANIPULATION
FASTA/FASTQ
FASTQ Quality Control
SAM/BAM
BED
VCF/BCF
Nanopore
Convert Formats
Lift-Over
COMMON GENOMICS TOOLS

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

JTech 2020 Scholarship: CSHL Biological Data Science

The JTech Foundation and Cold Spring Harbor Laboratory are pleased to announce the JTech 2020 Scholarship. JTech and CSHL will provide support to 10 outstanding graduate students in genomics and data sciences to attend the 2020 CSHL Biological Data Science Conference.

 + 

Application deadline is August 1, 2020 (Midnight EST). Applicants will be notified by August 8, 2020. Selected applicants will receive a stipend of \$2,500.

History

search datasets

Unnamed history
1 shown, 1 deleted
(empty)

2: UCSC Main on Human
an: dbSnp153Common (c
hr17:7,668,402-7,687,538)

jobs in green are complete, grey are queued, yellow are running, and jobs in red have failed or have errors

each job in the history has a name and a number, and can be rerun! (reproducibility)

⚠ All tools should be functioning normally with the exception of RNA STAR.

Tools



search tools



Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

Lift-Over

COMMON GENOMICS TOOLS

Create a genomic intervals

Chrom	Start	End	Name
chr17	7668781	7668782	rs17884306_up_1_chr17_766
chr17	7668834	7668835	rs200378797_up_1_chr17_766
chr17	7669037	7669038	rs1555523911_up_1_chr17_766
chr17	7669122	7669123	rs4968187_up_1_chr17_7669
chr17	7669909	7669910	rs6503048_up_1_chr17_7669
chr17	7670063	7670064	rs59098034_up_1_chr17_7670
chr17	7670995	7670996	rs17883852_up_1_chr17_7670
chr17	7671401	7671402	rs17879377_up_1_chr17_7671
chr17	7671455	7671456	rs1641549_up_1_chr17_76714
chr17	7671459	7671460	rs1641548_up_1_chr17_76714
chr17	7671544	7671545	rs9891744_up_1_chr17_76715
chr17	7671616	7671617	rs858528_up_1_chr17_76716
chr17	7672244	7672245	rs1642793_up_1_chr17_76722
chr17	7672413	7672414	rs12949655_up_1_chr17_76722
chr17	7672956	7672957	rs60699444_up_1_chr17_76729
chr17	7673028	7673029	rs75732100_up_1_chr17_76730
chr17	7673181	7673182	rs77697176_up_1_chr17_76731
chr17	7673521	7673522	rs1800899_up_1_chr17_76731
chr17	7674087	7674088	rs12951053_up_1_chr17_76740
chr17	7674107	7674108	rs12947788_up_1_chr17_76740
chr17	7674637	7674638	rs8069054_up_1_chr17_76740
chr17	7674795	7674796	rs1625895_up_1_chr17_76741
chr17	7675325	7675326	rs2909430_up_1_chr17_76753
chr17	7675359	7675360	rs9895829_up_1_chr17_76753
chr17	7675517	7675518	rs1794287_up_1_chr17_76753

History



search datasets



Unnamed history

2 shown, 1 deleted

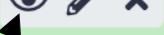
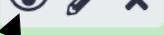
(empty)



3: UCSC Main on Human: ncbiRefSeq (chr17:7,68,402-7,687,538)

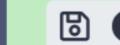


2: UCSC Main on Human: dbSnp153Common (chr17:7,668,402-7,687,538)



70 regions

format: bed, database: hg38



display in IGB View

display with IGV local Human hg38

display at UCSC main

1. Chrom	2. Start	3. End	4. Name
chr17	7668781	7668782	rs17884306_up_1_chr17_766
chr17	7668834	7668835	rs200378797_up_1_chr17_766
chr17	7669037	7669038	rs1555523911_up_1_chr17_766

click on the eye to view results



how many of these SNPs are in coding exons?

steps:

- get annotation for TP53 exons in hg38
- intersect with SNPs
- count



Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

Projects

Help

About Us

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. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: Mammal genome: Human assembly: Dec. 2013 (GRCh38/hg38)

group: Genes and Gene Predictions track: NCBI RefSeq add custom tracks track browser

table: RefSeq All (ncbiRefSeq) describe table schema

region: genome position chr17:7,668,402-7,687,538 lookup define regions

identifiers (names/acccessions): paste list upload list

filter: create

subtrack merge: create

intersection: create

correlation: create

output format: BED - browser extensible data Send output to Galaxy GREAT

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

get output summary/statistics

To reset all user cart settings (including custom tracks), [click here](#).

human
hg38
genes & gene
predictions
NCBI RefSeq
region == TP53
BED format

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

Output ncbiRefSeq as BED

Include custom track header:

name= tb_ncbiRefSeq

description= table browser query on ncbiRefSeq

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

each coding exon will be a separate entry

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

[Send query to Galaxy](#)

[Cancel](#)

⚠ All tools should be functioning normally with the exception of RNA STAR.

Tools



search tools



vCF/BCF

Nanopore

Convert Formats

Lift-Over

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Wiggle-to-Interval converter

Aggregate datapoints Append average, min, max of datapoints interval

Gene BED To Exon/Intron/Codon BED

expander

Get flanks returns flanking region/s for every gene

Subtract the intervals of two datasets

Join the intervals of two datasets side-by-side

Intersect the intervals of two datasets

Coverage of a set of intervals on second set of intervals

Complement intervals of a dataset

Chrom	Start	End	Name
chr17	7668781	7668782	rs17884306_up_1_chr17_766
chr17	7668834	7668835	rs200378797_up_1_chr17_766
chr17	7669037	7669038	rs1555523911_up_1_chr17_766
chr17	7669122	7669123	rs4968187_up_1_chr17_7669
chr17	7669909	7669910	rs6503048_up_1_chr17_7669
chr17	7670063	7670064	rs59098034_up_1_chr17_7670
chr17	7670995	7670996	rs17883852_up_1_chr17_7670
chr17	7671401	7671402	rs17879377_up_1_chr17_7671
chr17	7671455	7671456	rs1641549_up_1_chr17_76714
chr17	7671459	7671460	rs1641548_up_1_chr17_76714
chr17	7671514	7671515	rs2901711_up_1_chr17_76715

we want to know how many intervals in the SNP data overlap intervals in the exon table

1.Chrom	2.Start	3.End	4.Name
chr17	7672956	7672957	rs60699444_up_1_chr17_7672
chr17	7673028	7673029	rs75732100_up_1_chr17_7673
chr17	7673181	7673182	rs77697176_up_1_chr17_7673
chr17	7673521	7673522	rs1800899_up_1_chr17_7673
chr17	7674087	7674088	rs12951053_up_1_chr17_7674
chr17	7674107	7674108	rs12947788_up_1_chr17_7674
chr17	7674637	7674638	rs8069054_up_1_chr17_7674
chr17	7674795	7674796	rs1625895_up_1_chr17_7674
chr17	7675325	7675326	rs2909430_up_1_chr17_7675
chr17	7675359	7675360	rs9895829_up_1_chr17_7675
chr17	7675517	7675518	rs1794287_up_1_chr17_7675

History



search datasets

Unnamed history

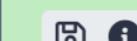
2 shown, 1 deleted

12.15 KB

3: UCSC Main on Human:
ncbiRefSeq (chr17:7,668,
402-7,687,538)

2: UCSC Main on Human:
chr1
38)

hg38



display in IGB View

display with IGV local Human hg38

display at UCSC main

1.Chrom	2.Start	3.End	4.Name
chr17	7668781	7668782	rs17884306_up_1_ch
chr17	7668834	7668835	rs200378797_up_1_c
chr17	7669037	7669038	rs1555523911_up_1_
chr17	7669122	7669123	rs4968187_up_1_ch
chr17	7669909	7669910	rs6503048_up_1_ch

⚠ All tools should be functioning normally with the exception of RNA STAR.

Tools



search tools

vCF/BCF

Nanopore

Convert Formats

Lift-Over

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Wiggle-to-Interval converter

Aggregate datapoints Appends the average, min, max of datapoints per interval

Gene BED To Exon/Intron/Codon BED expander

Get flanks returns flanking region/s for every gene

Subtract the intervals of two datasets

Join the intervals of two datasets side-by-side

Intersect the intervals of two datasets

Coverage of a set of intervals on second set of intervals

Complement intervals of a dataset

Intersect the intervals of two datasets (Galaxy Version 1.0.0)

▼ Options

Return

Overlapping Intervals

(see figure below)

of



2: UCSC Main on Hu...



First dataset

that intersect



3: UCSC Main on Hu...



Second dataset

for at least

1

(bp)

✓ Execute

ℹ **TIP:** If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

History



search datasets

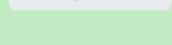
Unnamed history

2 shown, 1 deleted

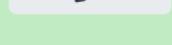
12.15 KB



3: UCSC Main on Human:
ncbiRefSeq (chr17:7,668,
402-7,687,538)

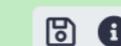
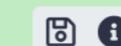


2: UCSC Main on Human:
dbSnp153Common (chr1
7:7,668,402-7,687,538)



70 regions

format: **bed**, database: **hg38**



display in IGB View

display with IGV local Human hg38

display at UCSC main

1.Chrom	2.Start	3.End	4.Name
chr17	7668781	7668782	rs17884306_up_1_ch
chr17	7668834	7668835	rs200378797_up_1_c
chr17	7669037	7669038	rs1555523911_up_1_
chr17	7669122	7669123	rs4968187_up_1_chr
chr17	7669909	7669910	rs6503048_up_1_chr



⚠ All tools should be functioning normally with the exception of RNA STAR.

Tools[search tools](#)[vcf/dvcf](#)[Nanopore](#)[Convert Formats](#)[Lift-Over](#)**COMMON GENOMICS TOOLS**[Operate on Genomic Intervals](#)[Wiggle-to-Interval converter](#)

Aggregate datapoints Appends the average, min, max of datapoints per interval

Gene BED To Exon/Intron/Codon BED

expander

Get flanks returns flanking region/s for every gene

Subtract the intervals of two datasets

Join the intervals of two datasets side-by-side

Intersect the intervals of two datasets

Coverage of a set of intervals on second set of intervals

Complement intervals of a dataset

Chrom	Start	End	Name
chr17	7676152	7676153	rs1042522_up_1_chr17_7676152

History[search datasets](#)**Unnamed history**

3 shown, 1 deleted

12.2 KB



4: Intersect on data 3 and data 2

1 region

format: **bed**, database: **hg38**[display in IGB View](#)[display with IGV local Human hg38](#)[display at UCSC main](#)

1. Chrom 2. Start 3. End 4. Name

chr17	7676152	7676153	rs1042522_up_1_chr17_7676152
-------	---------	---------	------------------------------

3. UCSC Main on Human:

ncbiRefSeq (chr17:7,668,402-7,687,538)

134 regions

format: **bed**, database: **hg38**[display in IGB View](#)[display with IGV local Human hg38](#)

when this finishes, click View to see the results

Galaxy workflows

- your history can be saved (if you are logged in) and you can convert the steps to a workflow
- workflows can be shared among users and can be rerun, promoting reproducibility

exercises

1. Look at the YHR203C gene in *Saccharomyces cerevisiae*. What's different about it? hint: not many yeast genes have introns! In fact, roughly how many introns are in *S. cerevisiae*? (Genome Browser, Table Browser)
2. How many NCBI RefSeq coding exons on human chromosome 19 overlap repeats? (Table Browser, Galaxy)
3. For all genes on human chromosome 19, find the mouse and chimpanzee orthologs and the % identity of the human gene to those orthologs. What trends do you notice? (BioMart)