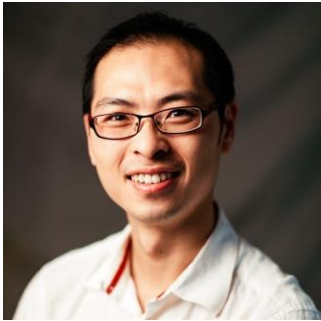


Open Data Resources for Human Genomics Research



A/Prof Jason Wong
Head, Computational Cancer Genomics
School of Biomedical Sciences
University of Hong Kong



Dr Rebecca Poulos
NHMRC Early Career Fellow
Children's Medical Research Institute
University of Sydney

Part One

Human genomics data resources

1:30pm – 3pm

- UCSC Genome Browser + Exercise
- gnomAd + Exercise
- GTEx + Exercise
- ENCODE database + Exercise
- Exercise: Putting it all together

Part Two

Cancer genomics data resources

3:30pm – 5pm

- Brief introduction to TCGA
- cBioPortal
- Genomic Data Commons
- Xena Browser
- Cancer Genomics Cloud

Part One

Human genomics data resources

1:30pm – 3pm

- UCSC Genome Browser

Visualise and download genomic datasets.

- gnomAd

Find data on single nucleotide polymorphisms (SNPs)

- GTEx

Get data for tissue-specific gene expression and regulation

- ENCODE database

Identify functional elements in the human genome

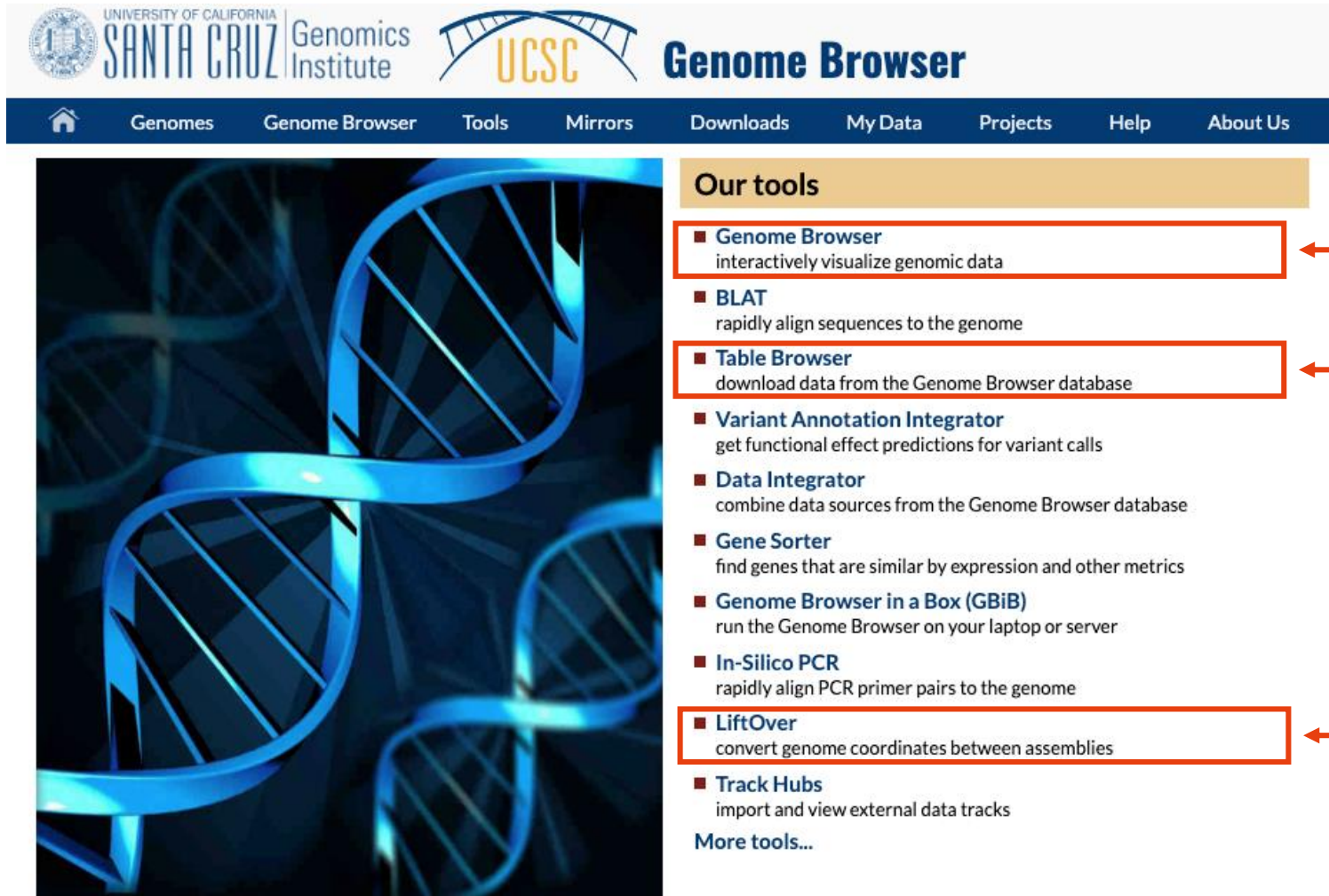
Each section will be accompanied by an exercise to help you test your skills, plus an exercise at the end to put all your knowledge together.

UCSC Genome Browser

Visualise and download genomic datasets.

UCSC Genome Browser

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



The image shows the UCSC Genome Browser homepage. At the top, there is a header with the University of California Santa Cruz Genomics Institute logo and the UCSC logo. Below the header is a navigation bar with links: Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. The main content area is divided into two sections. On the left, there is a large image of a DNA double helix. On the right, there is a section titled 'Our tools' which lists various tools and their functions. The tools listed are: Genome Browser (interactively visualize genomic data), 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), Gene Sorter (find genes that are similar by expression and other metrics), 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), and Track Hubs (import and view external data tracks). There is a 'More tools...' link at the bottom of the list.

Our tools

- **Genome Browser**
interactively visualize genomic data
- **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
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **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

[More tools...](#)

Visualise genomic data

Download genomic data

Convert genomic data from one reference genome to another

UCSC Genome Browser

[link](#)

Genome Browser

Gene (TP53)

Conservation

SNPs

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

Where you are located

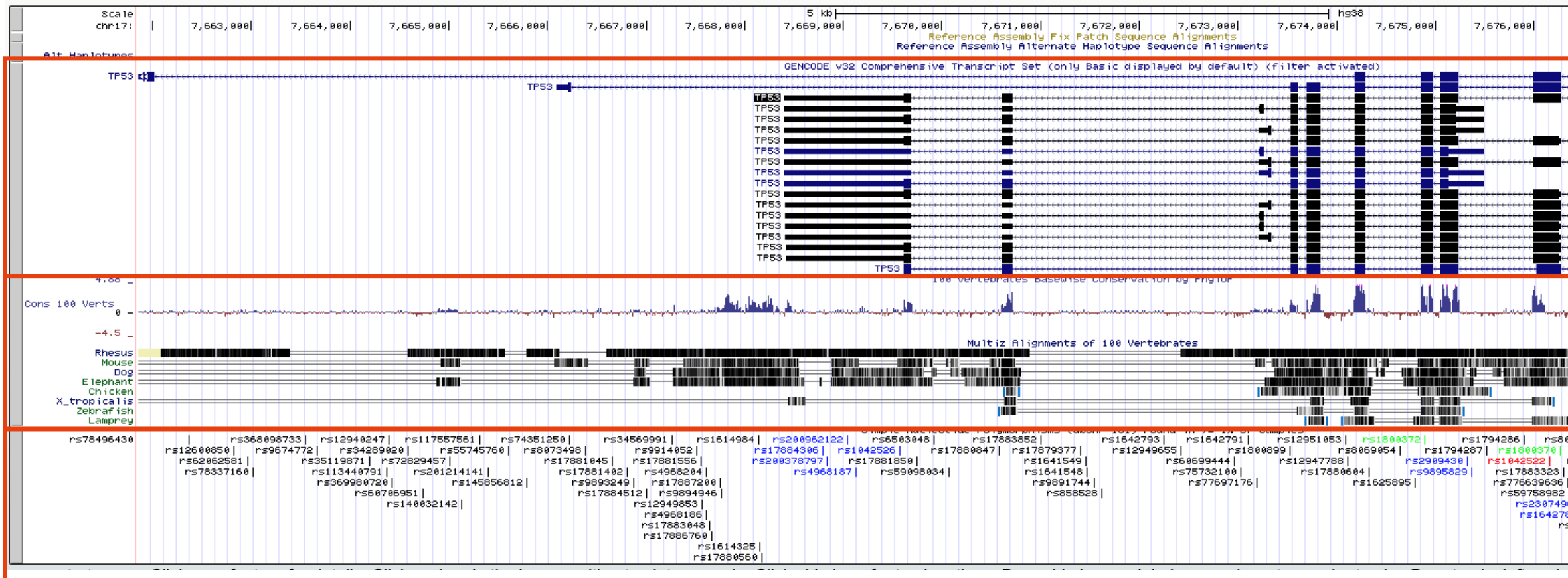
chr17:7,661,853-7,680,989 19,137 bp.

enter position, gene symbol, HGVS or search terms

Search

go

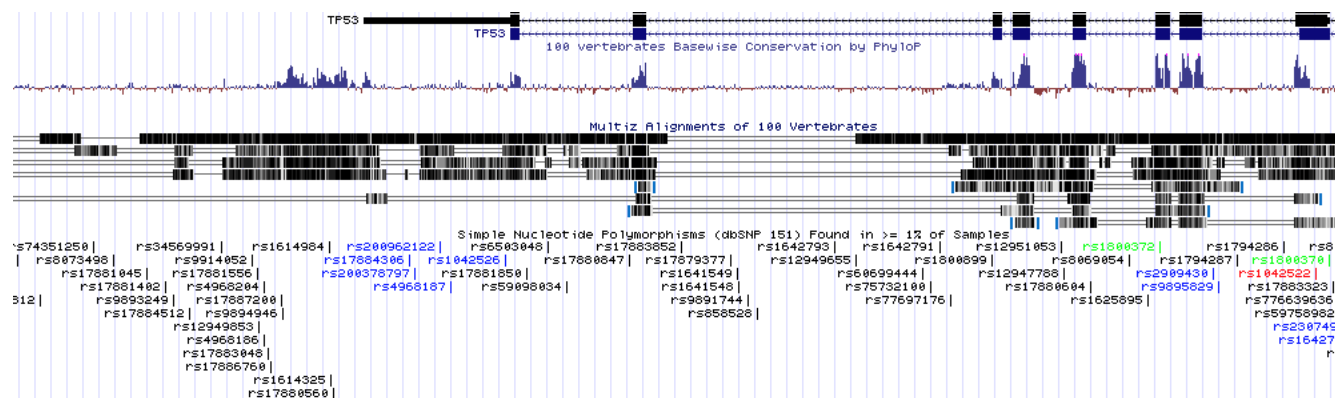
chr17 (p13.1) 17p13.3 17p13.2 17p13.1 17p12 17p11.2 17q11.2 17q12 17q21.2 17q21.31 q21.32 q21.33 17q22 17q23.2 23.3 24.1 17q24.2



Genome Browser



Genome Browser



Use 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 zoom in.

Add other tracks

Add custom tracks from your own data

Select common tracks to load

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.

Mapping and Sequencing

Base Position <input type="button" value="dense"/>	P12 Fix Patches <input type="button" value="pack"/>	P12 Alt Haplotypes <input type="button" value="dense"/>	P12 Assembly <input type="button" value="hide"/>	Centromeres <input type="button" value="hide"/>	P12 Chromosome Band <input type="button" value="hide"/>
Clone Ends <input type="button" value="hide"/>	P12 FISH Clones <input type="button" value="hide"/>	P12 Gap <input type="button" value="hide"/>	P12 GC Percent <input type="button" value="hide"/>	GRC Contigs <input type="button" value="hide"/>	GRC Incident <input type="button" value="hide"/>
Hg19 Diff <input type="button" value="hide"/>	P12 INSDC <input type="button" value="hide"/>	LRG Regions <input type="button" value="hide"/>	Mappability... <input type="button" value="hide"/>	P12 RefSeq Acc <input type="button" value="hide"/>	Restr Enzymes <input type="button" value="hide"/>
Scaffolds <input type="button" value="hide"/>	Short Match <input type="button" value="hide"/>	STS Markers <input type="button" value="hide"/>			

Genes and Gene Predictions

P12 GENCODE v32 <input type="button" value="pack"/>	NCBI RefSeq <input type="button" value="pack"/>	P12 Other RefSeq <input type="button" value="hide"/>	P12 Updated All GENCODE... <input type="button" value="hide"/>	P12 AUGUSTUS <input type="button" value="hide"/>	CCDS <input type="button" value="hide"/>
CRISPR Targets <input type="button" value="hide"/>	Geneid Genes <input type="button" value="hide"/>	P12 Genscan Genes <input type="button" value="hide"/>	P12 IKMC Genes Mapped <input type="button" value="hide"/>	LRG Transcripts <input type="button" value="hide"/>	MANE select v0.6 <input type="button" value="hide"/>
P12 MGC Genes <input type="button" value="hide"/>	Non-coding RNA... <input type="button" value="hide"/>	Old UCSC Genes <input type="button" value="hide"/>	P12 ORFeome Clones <input type="button" value="hide"/>	P12 Pfam in UCSC Gene <input type="button" value="hide"/>	RetroGenes V9 <input type="button" value="hide"/>

Genome Browser

Save and share your data

- **Make an account and save your sessions**
You can share these with collaborators

**View other
reference
genomes**

UCSC Genome Browser

[link](#)

Table Browser

Download genomic datasets



[Home](#) [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 DNase-seq data. For more information, see the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation. You can also examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for further analysis. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: **genome:** **assembly:**

group: **track:**

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: ☐ [Galaxy](#) ☐ [GREAT](#) ☐ [GenomeSpace](#)

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

file type returned: ☒ plain text ☐ gzip compressed

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

UCSC Genome Browser

[link](#)

LiftOver

Convert data from one reference genome to another

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

Lift Genome Annotations

This tool converts genome coordinates and genome annotation files between assemblies. The input data can be pasted into the text box, or uploaded from a file. If a pair of assemblies cannot be selected from the pull-down menus, a direct lift between them is unavailable. However, a sequential lift may be possible. Example: lift from Mouse, May 2004, to Mouse, Feb. 2006, and then from Mouse, Feb. 2006 to Mouse, July 2007 to achieve a lift from mm5 to mm9.

Original Genome:
Human

Original Assembly:
Dec. 2013 (GRCh38/hg38)

New Genome:
Human

New Assembly:
Feb. 2009 (GRCh37/hg19)

Minimum ratio of bases that must remap: 0.95

BED 4 to BED 6 Options

Allow multiple output regions: ☐

Minimum hit size in query: 0

Minimum chain size in target: 0

BED 12 Options

Min ratio of alignment blocks or exons that must map: 1

If thickStart/thickEnd is not mapped, use the closest mapped base: ☐

Paste in data ([BED](#) or chrN:start-end formats):

Submit
Clear

Or upload data from a file ([BED](#) or chrN:start-end in plain text format):

Choose file No file chosen

Submit File

Tools

Mirrors

Downloads

Blat

Table Browser

Variant Annotation Integrator

Data Integrator

Gene Interactions

Gene Sorter

Genome Graphs

In-Silico PCR

LiftOver

VisiGene

Other Utilities

UCSC Genome Browser

Exercise...

What is the location and nucleotide of the most frequent mutation arising in the KRAS gene in the TCGA cancer cohort?

Chromosome 12 at 25,245,350. C > T

Challenge...

What amino acid is present at this location in mouse?

(HINT: look at Conservation)

G / Glycine

Genome Aggregation Database (gnomAd)

Find data on single nucleotide polymorphisms (SNPs)

gnomAD v2.1.1 Search

gnomAD v3 released! 71,702 genomes aligned on GRCh38.

gnomAD

genome aggregation database

gnomAD v2.1.1 Search by gene, region, or variant

Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)

The [Genome Aggregation Database](#) (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community.

The v2 data set (GRCh37/hg19) provided on this website spans 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies. The v3 data set (GRCh38) spans 71,702 genomes, selected as in v2. The gnomAD Principal Investigators and groups that have contributed data to the current release are listed [here](#).

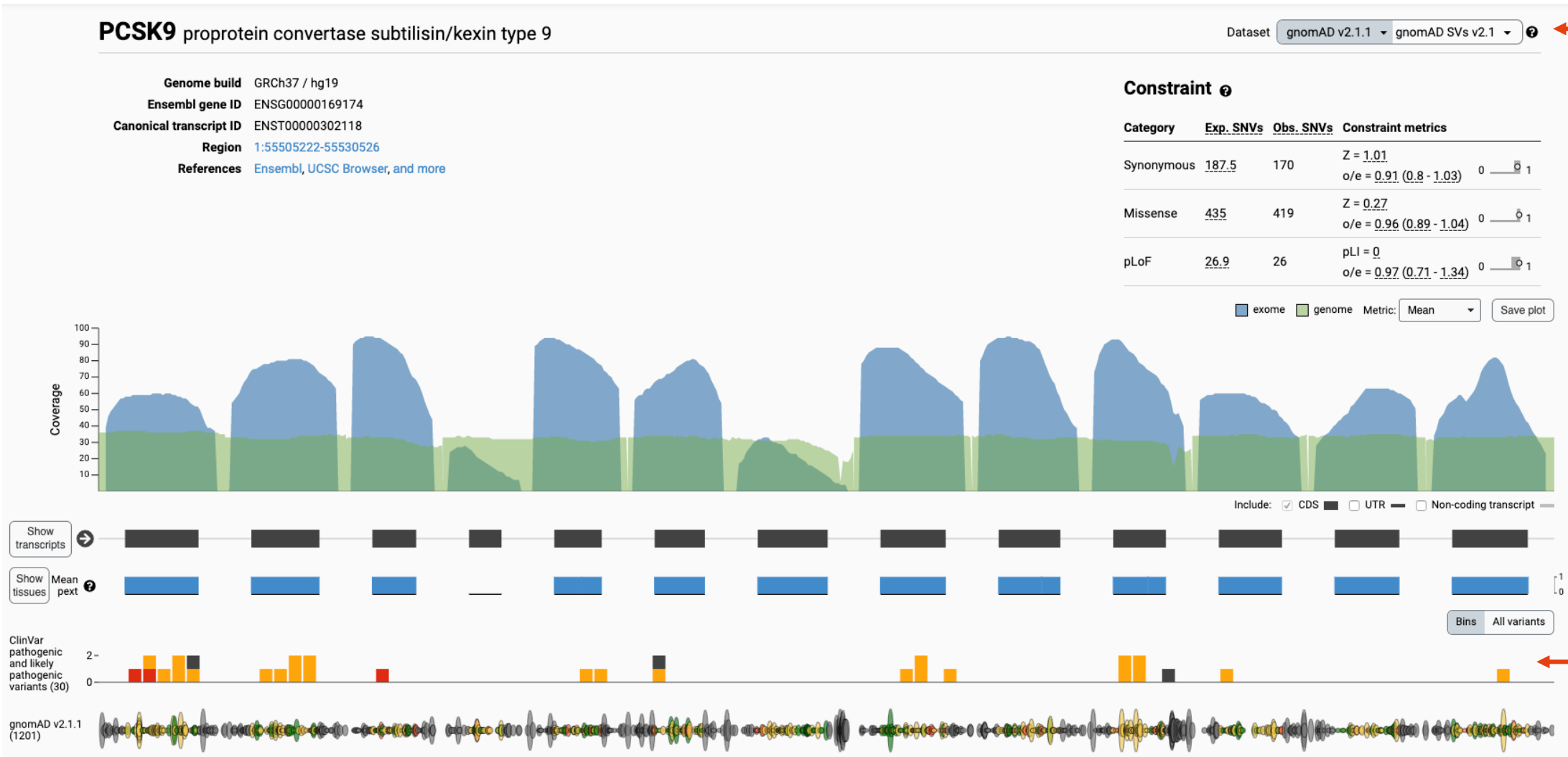
All data here are released for the benefit of the wider biomedical community, without restriction on use - see the terms of use [here](#). Sign up for our mailing list for future release announcements [here](#).

Search for a gene or variant

gnomAD

[link](#)

Swap to
structural
variants



Find
pathogenic
variants

Insertion: 1-55516888-G-GA (GRCh37)

Dataset gnomAD v2.1.1 ⓘ

View allele frequencies

Filter	Exomes	Genomes	Total
	Pass	Pass	
Allele Count	121	156	277
Allele Number	128230	31382	159612
Allele Frequency	0.0009436	0.004971	0.001735
Popmax Filtering AF (95% confidence)	0.01338	0.01542	
Number of homozygotes	2	0	2

This variant is multiallelic. Other alt alleles are:

- 1-55516888-G-A

Annotations

This variant falls on 4 transcripts in 1 gene.

frameshift

- PCSK9
 - ENST00000452118
 - HGVSp: p.Gly183GlufsTer23
 - pLoF: ● High-confidence

intron

- PCSK9
 - ENST00000302118 *
 - ENST00000543384

non coding transcript exon

- PCSK9
 - ENST00000490692

References

- dbSNP (rs527413419)
- UCSC *View in UCSC*

Report

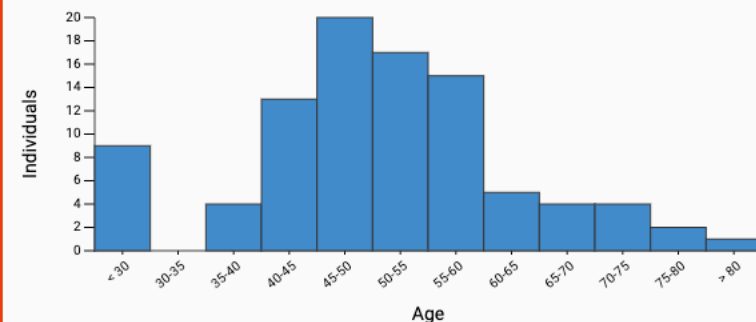
- Report this variant
- Request additional information

See breakdown of population frequencies

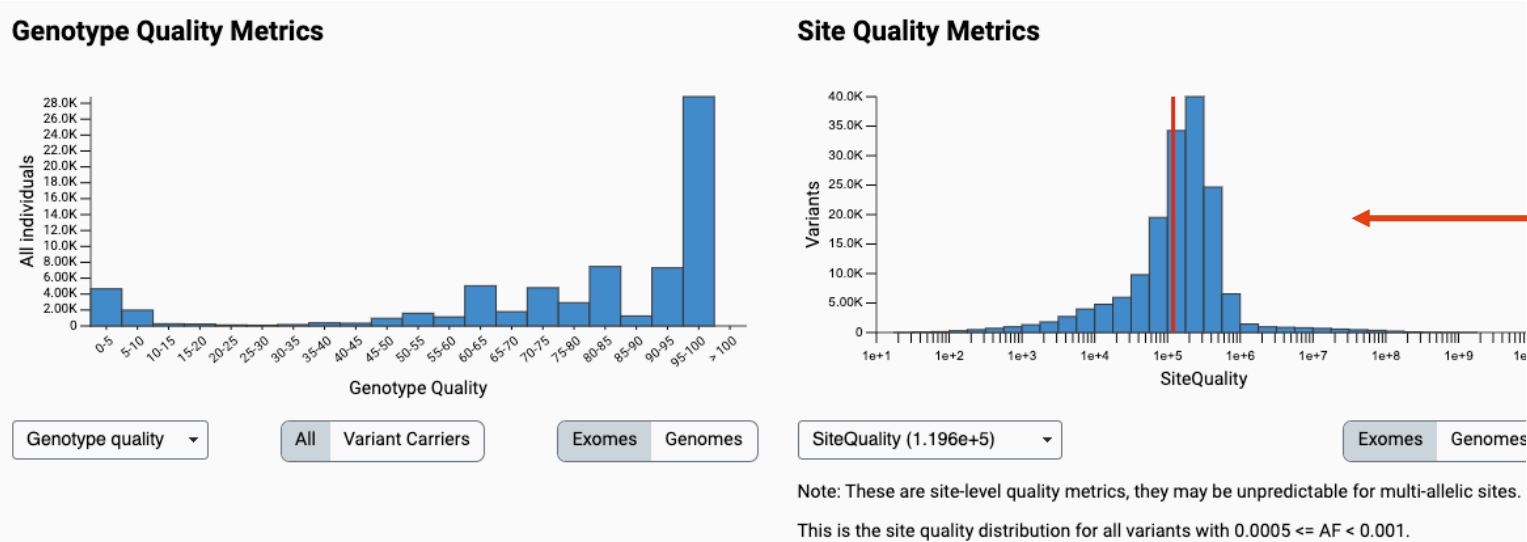
Population Frequencies ⓘ

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
▶ African	251	14784	2	0.01698
▶ Latino	18	25174	0	0.0007150
▶ Other	3	5074	0	0.0005912
▶ European (non-Finnish)	5	62976	0	0.00007940
▶ Ashkenazi Jewish	0	8378	0	0.000
▶ East Asian	0	11986	0	0.000

Age Distribution



See age breakdown



Quality metrics

Read Data

This interactive [IGV.js](#) visualization shows reads that went into calling this variant.

Note: These are reassembled reads produced by [GATK HaplotypeCaller --bamOutput](#), so they accurately represent what HaplotypeCaller was seeing when it called this variant.



Visualise raw data

gnomAD

Exercise...

What is the size of the structural variant in the BRAF gene with the highest allele frequency?

515 bp

Challenge...

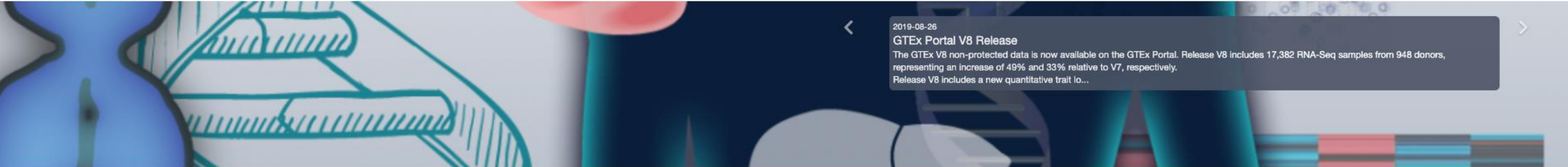
In what population (ie ethnicity) does this variant have the highest allele frequency?

East Asian

Genotype-Tissue Expression (GTEx) Portal

Get data for tissue-specific gene expression and regulation

We are currently working on optimizing our QTL calculator APIs. We have temporarily removed access to the eQTL Calculator page and eQTL Dashboard. During this time, the QTL violin plots will also not work. We apologize for any inconvenience this may cause. Please contact us if you have any questions.



2019-08-26
GTEEx Portal V8 Release
The GTEEx V8 non-protected data is now available on the GTEEx Portal. Release V8 includes 17,382 RNA-Seq samples from 948 donors, representing an increase of 49% and 33% relative to V7, respectively. Release V8 includes a new quantitative trait lo...

Resource Overview

Current Release (V8)

[Tissue & Sample Statistics](#)
[Tissue Sampling Info \(Anatomogram\)](#)
[Access & Download Data](#)
[Release History](#)
[How to cite GTEEx?](#)

Non-diseases tissue sites

The Genotype-Tissue Expression (GTEx) project is an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation. Samples were collected from 54 non-diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq. Remaining samples are available from the GTEx Biobank. The GTEx Portal provides open access to data including gene expression, QTLs, and histology images.

News and Events

2019-08-26 [GTEx Portal V8 Release](#)
2019-07-24 [GTEx V8 data release](#)

Explore GTEx

Browse

- [By gene ID](#)
- [By variant or rs ID](#)
- [By Tissue](#)
- [Histology Image Viewer](#)

Expression

- [Multi-Gene Query](#)
- [Top 50 Expressed Genes](#)
- [Transcript Browser](#)

- Browse and search all data by gene
- Browse and search all data by variant
- Browse and search all data by tissue
- Browse and search GTEx histology images
- Browse and search expression by gene and tissue
- Visualize the top 50 expressed genes in each tissue
- Visualize transcript expression and isoform structures

Tissue Page

Colon - Sigmoid

Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

Colon - Sigmoid (n=318)

Change Tissue

Tissue Information

Colon - Sigmoid	Counts
Total samples	373
Total samples with donor genotype	318
Number of significant eGenes ⓘ	10550
Number of significant sGenes ⓘ	3269

Top 100 Expressed Genes in Colon - Sigmoid

[Copy](#) [CSV](#) [Include Mitochondrial Genes](#) [Exclude Mitochondrial Genes](#)

Search: Show entries

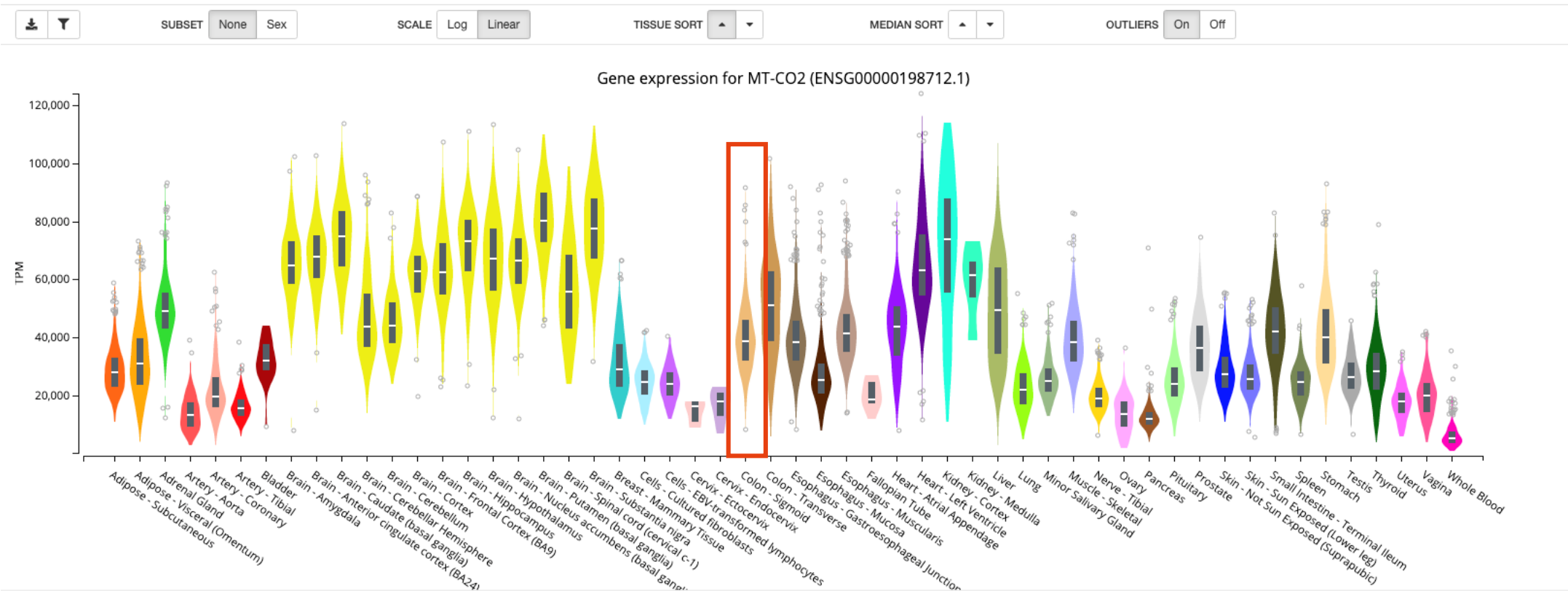
Gencode Id	Gene Symbol	Median TPM
ENSG00000198712.1	MT-CO2	38803.7
ENSG00000198886.2	MT-ND4	36183.8
ENSG00000198899.2	MT-ATP6	35550.3
ENSG00000198938.2	MT-CO3	33373.4
ENSG00000198804.2	MT-CO1	27547.2

GTEx Portal

Gene expression for MT-CO2 (ENSG00000198712.1)

Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

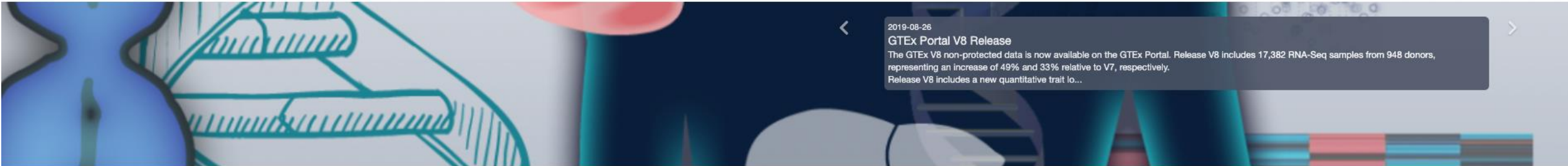
Data processing and normalization ⓘ



GTEEx Portal



We are currently working on optimizing our QTL calculator APIs. We have temporarily removed access to the eQTL Calculator page and eQTL Dashboard. During this time, the QTL violin plots will also not work. We apologize for any inconvenience this may cause. Please contact us if you have any questions.



Resource Overview

Current Release (V8)

[Tissue & Sample Statistics](#)
[Tissue Sampling Info \(Anatomogram\)](#)
[Access & Download Data](#)
[Release History](#)
[How to cite GTEEx?](#)

The Genotype-Tissue Expression (GTEEx) project is an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation. Samples were collected from 54 non-diseased tissue sites across nearly 1000 individuals, primarily for molecular assays including WGS, WES, and RNA-Seq. Remaining samples are available from the GTEEx Biobank. The GTEEx Portal provides open access to data including gene expression, QTLs, and histology images.

News and Events

2019-08-26 [GTEEx Portal V8 Release](#)
2019-07-24 [GTEEx V8 data release](#)

Explore GTEEx

Browse

- [By gene ID](#)
- [By variant or rs ID](#)
- [By Tissue](#)
- [Histology Image Viewer](#)

Expression

- [Multi-Gene Query](#)
- [Top 50 Expressed Genes](#)
- [Transcript Browser](#)

Browse and search all data by gene

Browse and search all data by variant

Browse and search all data by tissue

Browse and search GTEEx histology images

Browse and search expression by gene and tissue

Visualize the top 50 expressed genes in each tissue

Visualize transcript expression and isoform structures

Gene Symbol	Gencode ID	Entrez Gene ID	Location	Gene Description	Actions
KLK3	ENSG00000142515.14	354	chr19:50854915-50860764:+	kallikrein related peptidase 3 [Source:HGNC Symbol;Acc:HGNC:6364]	<div>IGV eQTL Browser, Ensembl, UCSC</div>

Showing 1 to 1 of 1 entries

FirstPrevious1NextLast

Gene expression for KLK3 (ENSG00000142515.14)

Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)
Data processing and normalization

SUBSETNoneSex

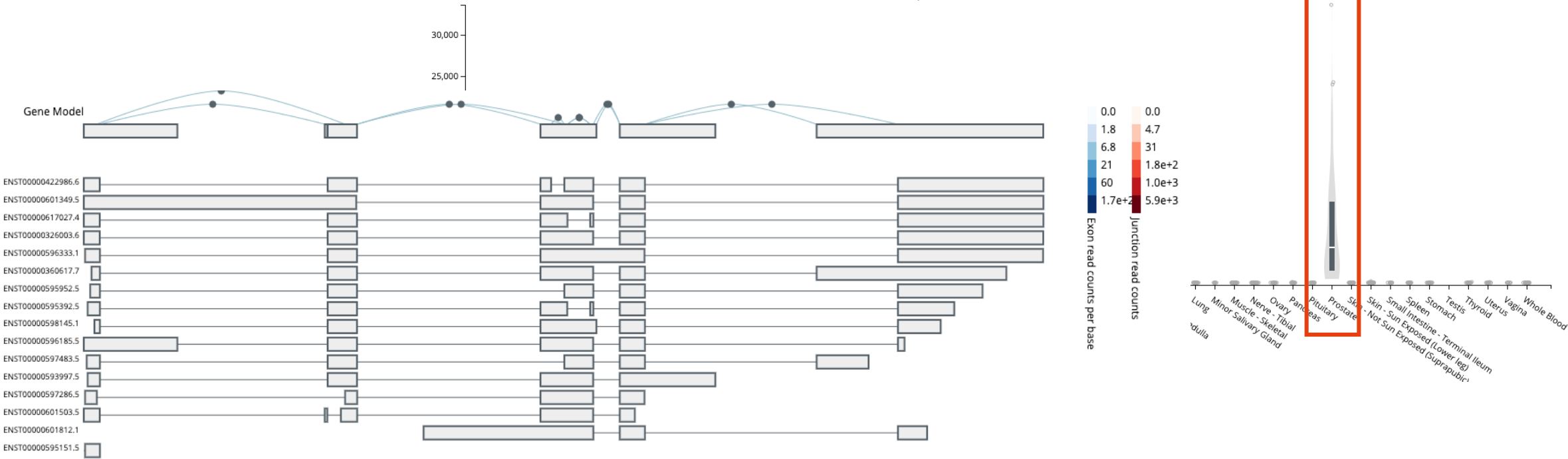
SCALELogLinear

TISSUE SORT

MEDIAN SORT

OUTLIERSOnOff

Gene expression for KLK3 (ENSG00000142515.14)



GTEx Portal

Exercise...

Which cell type has the highest median expression of the PTEN gene?

EBV-transformed lymphocytes

Challenge...

Which PTEN exon has the highest median read count per base in this cell type?

Exon 8

Encyclopedia of DNA Elements (ENCODE)

Identify functional elements in the human genome

Search bar

Data Type

Dataset

Experiment

Biosample


BiosampleType

2

2

1

1

Audit category: 

extremely low read depth

1

Showing 4 of 4 results

RNA microarray of COLO829

Homo sapiens COLO829

Lab: Gregory Crawford, Duke

Project: ENCODE

Experiment

ENCSR000BFC

released

DNase-seq of COLO829

Homo sapiens COLO829

Lab: Gregory Crawford, Duke

Project: ENCODE

Experiment

ENCSR000EIN

released

1

COLO829 (Homo sapiens, adult 45 year)

Type: cell line

Summary: Homo sapiens COLO829 cell line

Source: ATCC

Biosample

ENCBS172AAA

released


COLO829 (cell line)

Ontology ID: EFO:0002140

Biosample Type


Add all items to cart

{ }




CHILDREN'S
MEDICAL
RESEARCH
INSTITUTE

Jeans for Genes®



THE UNIVERSITY OF
SYDNEY



香港大學
THE UNIVERSITY OF HONG KONG

AMSI
BIOINFO SUMMER
A SYMPOSIUM IN BIOINFORMATICS

19

Data Type

Dataset	500
Experiment	369
File	159
Series	130
ReferenceGenome	126

Audit category:

extremely low read depth	27
missing control alignments	27
control extremely low read depth	4
inconsistent genetic modification reagent source and identifier	3

Audit category:

insufficient read depth	86
missing controlled_by	72
severe bottlenecking	34
poor library complexity	25
insufficient read length	13

Audit category:

mild to moderate bottlenecking	226
--------------------------------	-----

Showing 25 of 696 results

View All

Showing 25 of 696 results

TF ChIP-seq of gastroesophageal sphincter

Homo sapiens gastroesophageal sphincter male adult (54 years)

Target: CTCF

Lab: Michael Snyder, Stanford

Project: ENCODE

TF ChIP-seq of omental fat pad

Homo sapiens omental fat pad male adult (37 years)

Target: CTCF

Lab: Michael Snyder, Stanford

Project: ENCODE

TF ChIP-seq of limb

Target: CTCF

Lab: Richard Myers, HAIB

Project: ENCODE

TF ChIP-seq of gastroesophageal sphincter

Homo sapiens gastroesophageal sphincter male adult (37 years)

Add all items to cart

Experiment

ENCSR640BMC

revoked

1

Experiment

ENCSR742XZS

revoked

1

Experiment

ENCSR295PZC

revoked

Experiment

ENCSR393CZG

revoked

CHILDREN'S MEDICAL RESEARCH INSTITUTE
Jeans for Genes®

THE UNIVERSITY OF SYDNEY

香港大學
THE UNIVERSITY OF HONG KONG

AMSI
BIOINFO SUMMER 19
A SYMPOSIUM IN BIOINFORMATICS

ENCODE

[illegible]

Experiment Matrix

[Clear Filters](#)

Enter search terms to filter the experiments included in the matrix.

Assay type

DNA binding

9019

Transcription

3471

DNA accessibility

1095

RNA binding

699

DNA methylation

569

Assay title

TF ChIP-seq

3609

Histone ChIP-seq

3181

Control ChIP-seq

2229

DNase-seq

822

polyA plus RNA-seq

770

Selected filters:

released

15326

archived

1057

revoked

265

Project

ENCODE

9779

Roadmap

2154

modERN

1728

modENCODE

1095

GGR

553

RFA

ENCODE3

5992

ENCODE2

2703

Roadmap

2154

modERN

1728

modENCODE

1095

Genome assembly

Showing 15326 results

Download

Visualize

ASSAY

BIOSAMPLE

ENCODE

ENCODE

[Data](#)[Encyclopedia](#)[Materials & Methods](#)[Help](#)

Search...

Q

Experiment matrix

Search

Summary

Cloud Resources

AWS Open Data

Collections

ENCORE

ENTEx

SE Stem Cell Consortium

Reference epigenomes

Mouse development series

Search by region

Publications

ENCODE: Encyclopedia of DNA Elements

CH₃

CH₃CO

RNA polymerase

5C

ChIA-PET

Hi-C

DNase-seq

FAIRE-seq

ATAC-seq

ChIP-seq

WGBS

RRBS

methyl array

Computational predictions

RNA-seq

CLIP-seq

RIP-seq

Long-range regulatory elements

(enhancers, repressors/silencers, insulators)

Promoters

Genes

Transcripts

About ENCODE Project

Getting Started

Experiments

Search ENCODE portal

Q

About ENCODE Encyclopedia

candidate Cis-Regulatory Elements

Search for candidate Cis-Regulatory Elements

Hosted by SCREEN

Q

Human hg19

Q

Mouse mm10

Q

ENCODE

Data

Encyclopedia

Materials & Methods

Help

Search...

Reference epigenome search

Clear Filters

Assay title

Control ChIP-seq	319
Histone ChIP-seq	319
polyA plus RNA-seq	166
DNase-seq	151
WGBS	136

Status

Selected filters: released

released	319
archived	20

Genome assembly

GRCh38	227
hg19	227
mm10	92
mm9	21

Organism

Homo sapiens	227
Mus musculus	92

Target category

control	319
histone	319
narrow histone mark	319

Showing 25 of 319 results

View All

Reference epigenome series in upper lobe of left lung (*Homo sapiens*, adult, 37 year)

Targets: H3K4me1, H3K27me3, H3K4me3, H3K9me3, POLR2A, Control, EP300, H3K27ac, H3K36me3, CTCF

Lab: ENCODE Consortium

Project: ENCODE

Series ENCSR867OGI

released

Reference epigenome series in esophagus squamous epithelium (*Homo sapiens*, adult, 37 year)

Targets: H3K4me1, H3K27me3, H3K4me3, H3K9me3, POLR2A, Control, H3K27ac, H3K36me3, CTCF

Lab: ENCODE Consortium

Project: ENCODE

Series ENCSR284DRW

released

1

Reference epigenome series in tibial artery (*Homo sapiens*, adult, 37 year)

Targets: H3K4me1, H3K27me3, H3K4me3, H3K9me3, Control, H3K27ac, H3K36me3, CTCF

Lab: ENCODE Consortium

Project: ENCODE

Series ENCSR646ZPK

released

1

Reference epigenome series in Peyer's patch (*Homo sapiens*, adult, 37 year)

Targets: H3K4me1, H3K27me3, H3K4me3, H3K9me3, Control, H3K27ac, H3K36me3, CTCF

Lab: ENCODE Consortium

Project: ENCODE

Series ENCSR908HXO

released

1

Reference epigenome series in breast epithelium (*Homo sapiens*, adult, 51 year)

Targets: H3K4me1, H3K27me3, H3K4me3, H3K9me3, POLR2A, Control, POLR2AphosphoS5, H3K27ac, H3K36me3, CTCF

Lab: ENCODE Consortium

Project: ENCODE

Series ENCSR881FTZ

released

1

ENCODE

DataEncyclopediaMaterials & MethodsHelp

Search...

Datasets / Series / Reference Epigenome

Summary for reference epigenome series ENCSR881FTZ

1

Summary

Status:

● released

Description:

Reference epigenome for human breast epithelium (51 year old female)

Donor diversity:

single

Assay:

DNase-seq, RNA-seq, ATAC-seq, ChIP-seq

Biosample summary:

breast epithelium (*Homo sapiens*)

Attribution

Lab:

ENCODE Consortium

Award:

[U41HG006992](#) (J. Michael Cherry, Stanford)

Project:

ENCODE

External resources:

None submitted

Experiments in reference epigenome series ENCSR881FTZ

Add all items to cart

Accession	Assay	Target	Description	Lab	Status	Cart
ENCSR081OTO	ChIP-seq	H3K27ac		Bradley Bernstein, Broad	● released	
ENCSR134LLK	ChIP-seq	H3K27me3		Bradley Bernstein, Broad	● released	
ENCSR263XKR	ChIP-seq	H3K4me1		Bradley Bernstein, Broad	● released	
ENCSR304IVU	ChIP-seq	POLR2A	POLR2A ChIP-seq on human breast epithelium	Michael Snyder, Stanford	● released	
ENCSR400WEK	ChIP-seq	POLR2AphosphoS5	Chip-Seq on mammary gland	Richard Myers, HAIB	● released	
ENCSR438YPF	RNA-seq		The libraries contained in this Experiment come from tissue sub-sections of a 51 year old female's adipose tissue section obtained from GTEx. They are stranded PE101 Illumina Hi-Seq RNA-Seq libraries from rRNA-depleted Total RNA > 200 nucleotides in size.	Thomas Gingeras, CSHL	● released	
ENCSR568QQU	ChIP-seq	H3K4me3		Bradley Bernstein, Broad	● released	

CHILDREN'S MEDICAL RESEARCH INSTITUTE
Jeans for Genes®

THE UNIVERSITY OF SYDNEY

香港大學
THE UNIVERSITY OF HONG KONG

AMSI
BIOINFO SUMMER
A SYMPOSIUM IN BIOINFORMATICS
19

ENCODE

Exercise...

What is the experiment number (ie ENCXXXXXXXX) for WGBS in the K562 cell line?

ENCSR765JPC

Challenge...

What is the command you can use to download these data?

xargs -L 1 curl -O -L < files.txt

Exercise: Putting it all together

Exercise: Putting it all together

1. Go to the region chr5:112,838,930-112,838,940 in hg38. What gene spans this region?

APC

2. What is the methylation level (hint: use WGBS data) of this site in human sigmoid colon tissue? (You will need to use the ENCODE database).

Between 90-100%

3. How many samples in the TCGA cancer cohort are mutated at codon R114?

24

4. How many missense SNPs are present in this gene according to the gnomAD browser?

1,436

5. In which organ does this gene have the highest median expression? (Try to answer this question without leaving the UCSC genome browser).

Brain

See you back again at 3:30pm!

Part Two

Cancer genomics data resources

3:30pm – 5pm

- Brief introduction to TCGA
- cBioPortal
- Genomic Data Commons
- Xena Browser
- Cancer Genomics Cloud