Genomic Analysis with Genome Browsers

http://barc.wi.mit.edu/hot_topics/





Outline

- Genome browsers overview
- UCSC Genome Browser
 - Navigating: View your list of regions in the browser
 - Available tracks (eg. ENCODE, GTEx, etc.)
 - Table Browser to download and annotate genome features
 - Convert coordinates/features between genomes
 - Use Public Hub to display tracks hosted at non-UCSC servers.
 - Saving your session
- Ensembl Browser
 - Gene Tree
 - BioMart



Genome Browsing

- Facilitate genomic analysis in the context of genomic sequences, such as
 - Alignments (eg. conservation)
 - Experimental/Annotation Data (eg. TFBS)
- Commonly used browsers:
 > UCSC Genome Browser (UCSC)
 > EMBL-EBI Ensembl
 > NCBI Map Viewer

Cline, M.C, and Kent, J.W Understanding Genome Browsing Nat Biotech (2009) WHITEHEAD INSTITUTE

UCSC Genome Browser

- Created in 2001
- Browse almost 100 genomes (as of 2017)
- Web-based and open source platform
- Many datasets are hosted on UCSC
- Visualize your own genome-mapped datasets or custom data
 - save images as PDF/EPS for publication
- Easily download genome features with Table Browser





UCSC Genome Browser: Home Page



*UCSC Genome Browser currently doesn't host any plant genomes

WHITEHEAD INSTITUTE

Yeast

UCSC Genome Browser: genome.ucsc.edu Local Whitehead Mirror: membrane.wi.mit.edu





Choosing a Genome/Assembly



Whitehead Institute

Navigation

Move horizontally by 10%,50%,95%

	🕋 G	enomes	Genome Browser	Tools	Mirrors	Downloads	My Data	View	Help	About Us
		SC Gen ve <<<	ome Brows	er on l	Human	Dec. 2013	(GRCh3	88/hg3 ut 1.5x	8) Ass 3x 10x	embly
n ation	Ideog	chr9:133,23	38,935-133,289,71	6 50,782	bp. enter po	sition, gene symbo	ol or search tern	ns		_{go} Zoom In/Out
criptio nfigura		Scale	133 250 0	20 ki	> 	133 270 (hg38 280.000		
rack des track co	~	ABO ABO	GENCODE v2 Drag track and to new positio	2 Compi move	rehensive	Transcript Se	t (only Basi	c display k: item	red by di descri	efault) iption
lick: tı click:	move	e start 2.0 >	Click on a feat zoom in. Click down to reorde	ure for det side bars er tracks. [ails. Click or for track opt Drag tracks I	drag in the bas ions. Drag side eft or right to ne	se position tra bars or label ew position.	ack to s up or	rr < 2.	nove end
eft c ight	track search	default t	racks default order	hide all	add custon	n tracks track h esh	ubs configu	re multi-	region 🔶 r	everse resize
<u> </u>	Move st	 tart by)	(bases					Move	end by	X bases

Nove end by X bases



8

WHITEHEAD INSTITUTE

Navigation: Zoom-in with Drag-and-select

Drag/select at the top near chromosome position



Browser Tracks

lit ⊻iew Hi <u>s</u> tory <u>B</u> ookmark	s <u>T</u> ools <u>H</u> elp								
Human chr14:54416455-54	×								
genome. ucsc.edu /cgi-bin/h	gTracks?db=hg19&position=chr14%3A5441645	5-54423554&hgsid=4	19904723_ND3QHJdchuhQGQtX3	qg1U4jx8GBr					☆ 自 🕹 🎓 🤗
t Visited [] Getting Started	https://agrica.wi.mit.e								
-4.5 _				Multit a	lignments of 100 lies		d . deriver	and the second se	Least of Least 1
Rhesus Mouse									
Elephant Chicken X_tropicalis									
Zebrafish				Simple Nucleotide Polymor	phisms (dbSNP 142)	Found in >= 1% of Sampl	es.		
move start	Click on a feature for details.	Click or drag in	the base position track	to zoom in. Click sid	le bars for track o	ptions. Drag side b	ars or labels up or dow	vn to reorder tracks. Drag	move end
< 2.0 >	tracks left or right to new posit	tion.				· · · · · · · · · · · · · · · · · · ·			< 2.0 >
	track	k search defau	ult tracks default order	hide all add custo	om tracks track	hubs configure	reverse resize refi	iresh	
		collapse all	Use drop-down co Tracks with lots of iter	ontrols below and pre ms will automatically	ss refresh to alte be displaved in r	r tracks displayed. nore compact mode	expand all		
				Mapping and S	equencing		refresh		
		-		Genes and Gene	Predictions		refresh		
		UCSC	D (2) 0		0000	E 110			
		Genes	pack	hide -	hide -	hide -	hide -		
		pack 💌	,.				IKMC Genes		
		Exoniphy bide	GENCODE	Geneid Genes	Genscan Gene	s H-Inv 7.0	Mapped		
		Thue •	inde •				hide 💌		
		lincRNAs	LRG Transcripts	MGC Genes	N-SCAN	Old UCSC Gene	S Clones	click lin	ks for
		hide 💌	hide -	hide -	hide -	hide 💌	hide 🗸		
		Other RefSea	Pfam in UCSC Gene	Retroposed Genes	SGP Genes	SIB Genes	sno/miRNA	detail ir	formatio
		hide 🔻	hide 💌	hide 💌	hide -	hide 🝷	hide 💌		
		TransMap	tRNA Genes	UCSC Alt Events	<u>UniProt</u>	Vega Genes	Yale Pseudo60		
			nide		nide •	nide	nide •		
							Tellesi		
				mRNA and	d EST		retresh		
		<u>•</u>		Express	ion		refresh		
		•		Regulat	ion		refresh		
		•		Comparative	Genomics		refresh		
		•	Ne	andertal Assemb	ly and Analys	is	refresh		
		•	D	enisova Assembl	y and Analysis	s	refresh		
		•		Variati	on		refresh		
		•		Repea	ts		refresh		

The current genome build might have less tracks than the last version. For example: hg38 has considerably fewer tracks than hg19

WHITEHEAD INSTITUTE

Tracks*

- Mapping and Sequencing
- Genes and Gene Predictions
- Phenotype and Literature
- mRNA and EST
- Expression
- Regulation
- Comparative Genomics
- Variation
- Repeats

WHITEHEAD INSTITUTE

*Some tracks may not be available in your genome of interest



Track Settings and Description



The RefSeq Genes track shows known human protein-coding and non-protein-coding genes taken from the NCBI RN collection (RefSeq). The data underlying this track are updated weekly.

Please visit the <u>Feedback for Gene and Reference Sequences (RefSeq)</u> page to make suggestions, submit additions help concerning RefSeq records.

Display Conventions and Configuration

Item description



Position: chr9:133,255,666-133,275,189 Size: 19,524 Coding Exon Count: 8

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

BaRC

Configure: Image and Tracks

move start Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end
track search default tracks default order hide all manage custom tracks track hubs configure multi-region
r Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us
Configure Image
submit
image width: 1000 pixels
label area width: 30 characters
text size: 14 -
Display chromosome ideogram above main graphic
Show light blue vertical guidelines, or light red vertical window separators in multi-region view
Display labels to the left of items in tracks
Display description above each track
Show track controls under main graphic
Next/previous item navigation
Next/previous exon navigation
Show exon numbers
Enable highlight with drag-and-select (if unchecked, drag-and-select always zooms to selection)





Annotation Track Display Mode

Hide: the track is not displayed at all.

Dense: the track is displayed with all features collapsed into a single line.



Squish: the track is displayed with each annotation feature shown separately, but at 50% the height of full mode. Features are unlabeled.



Pack: the track is displayed with each annotation feature shown separately and labeled



Full: the track is displayed with each annotation feature on a separate line.





Gene Structure



Commonly used gene models

- GENCODE
- Ensembl
- Refseq





Multi-region: "Slice" to view or highlight exons



WHITEHEAD INSTITUTE

E

Demo and Exercise 1

• Finding your gene of interest

number of transcripts

Comparing RefSeq, Ensembl and GENCODE models

– EST evidence

Codons and amino acid





UCSC Tracks*

- Regulation ENCODE
- Expression GTEx
- Variation
- Comparative Genomics



ENCODE

Encyclopedia of DNA Elements

Identify functional genomic elements

DNA Elements: TF, histone modification, RNA binding protein, etc.

➤Various tissues and cell lines

 Extended to other model organisms (modENCODE)

UCSC has human/mouse: hg19 contains more data/tracks than hg38





ENCODE track: use settings to select specific samples/datasets



ENCODE track: configuration



WHITEHEAD INSTITUTE

ENCODE: More Info genome.ucsc.edu/ENCODE



WHITEHEAD INSTITUTE

ENCODE: search tracks

						,		
Search fo	r Tracks i	n the Human Fe	b. 2009 (C	GRCh37/hg	g19) Assembl	у		
Search	Adv	anced						
	Tra	ack Name: cor	ntains					
	and De	escription: cor	ntains					
	and Gr	oup:	is Any			-		
	and Da	ita Format:	is Any					-
								ENC
- +	and Antibod	y or target protein	• <mark>is amo</mark>	ong GATA1			Antibody	or target pro
- +	and Cell, tis	sue or DNA sample	+ <mark>is amo</mark>	ong Any		Cell,	tissue or D	NA sample
- +	and View - I	Peaks or Signals	- is amo					
		-		- 17 41 9				
search	clear c	ancel						
Jearch								
+ - Visit	oility Trac	k Name						
hide	K562	GATA1 K56	2 TFBS Ur	iform Peaks	s of GATA-1 fror	n ENCODE/L	JSC/Analy	sis 💌



Brov (hg1 Expe List Cell Mous UCSC Down



Demo and Exercise 2

• Use ENCODE data to look for TFBS, and gene expression (RNA-Seq) for your favorite gene





Expression*: GTEx

- Genotype-Tissue Expression (GTEx) project to study ~8600 tissues from ~600 (post-mortem) adults.
- Track is displayed by default in hg38, available in hg19 as well



*Other expression tracks available as well, eg. RNASeq from ENCODE and array data WHITEHEAD INSTITUTE



Variation

• Variants from several resources available: dbSNP, 1000 Genomes Project and ExAC



dbSNP track color-coded: Blue: UTR or ncRNA variant Red: coding non-synonymous Green: coding synonymous





Comparative Genomics: Conservation



left-click to configure track: select species

tiz Alignments pack Basewise Conservation (phyloP) full Element Conservation (phastCons) hide	Conserved Elements hide
Multiz Alignments Configuration	
	- Human
apecies selection.	Chimp
Primate .	- Orangutan
chimp gorilla crangular dibbon refesus	Gibbon
Crab-eating macaque baboon areen monkey marmoset squirrel monkey	Rhesus
E bushbaby	- Crab eating mac
	Green monkey
Euarchontoglires 📩 🚍	- Marmoset
Chinese tree shrew Squirrel I lesser Egyptian ierboa D prairie vole Chinese hamster	Squirrel monkey
🗉 golden hamster 😿 mouse 📰 rat 🔄 naked mole-rat 🖾 guinea pig	Chinese tree shi
🗈 chinchilla 🔹 brush-tailed rat 🖾 rabbit 👘 pika	Squirrel
	Lesser Egyptian
Lauraslatheria 📩 📥	Chinese hamster
🗖 pig 🔤 alpaca 📑 bactrian camel 🗖 dolphin 📑 killer whale	Golden hamster
Tibetan antelope Cow Sheep domestic goat horse	Mouse
🗉 white rhinoceros 🖾 cat 🛛 🗷 dog 🔅 ferret 💭 panda	- Hat
🗈 pacific watrus 📑 weddell seal 📑 black flying-fox 🖹 megabat 📄 david's myotis (bat)	Guinea pig
🗉 microbat 🔲 big brown bat 🗈 hedgehog 📄 shrew 📄 star-nosed mole	Chinchilla
	 Brush tailed rat Dabbit
Afrotheria 🗈 🖻	Pika
🗹 elephant 🗉 cape elephant shrew 🖾 manatee 🖾 cape golden mole 🖾 tenrec	Pig
aardvark	Alpaca
	Bactrian camel Dolobio
Mammal 🙏 🗖	Killer whale
🗖 armadillo 🗂 opossum 🗂 tasmanian devil 🗍 wallativ 🗂 platinus	Tibetan antelope
	Cow
Birds 🗶 🖻	Goat

PhyloP basewise conservation: rangesFrom -20 (least conserved) to7.532 (most conserved); other optionsinclude PhastCons and GERP

- Vertebrate Multiz Alignment using 100 species
- Fly 27-way alignment



Add custom track





Whitehead Institute

UCSC Genome Browser: Tools

- Blat
- Table Browser
- Gene Sorter
- LiftOver







Tools: Table Browser

<u>File E</u> dit <u>V</u> iew Hi <u>s</u> tory J	<u>B</u> ookmarks <u>T</u> ools <u>H</u> elp								_ • ×	
Table Browser	× +									
enome.ucsc.edu	/cgi-bin/hgTables?hgsid=419904	723_ND3QHJdchuhQGQtX3qg1U4jx8GBr				⊽ C'	Q Search	☆ 自 🤳	^	
Most Visited 🛄 Getting	Started https://agrica.wi.mi	t.e							<u>^</u>	
n	Genomes (Genome Browser	Iools Mirro	rs Downloads	My Data	Help	About Us			
Table Bro	Table Browser									
Use this prisequence of User's Guide features are your set that to the Creater from the Second for the second filter: Creater filter: Creater filter: Creater for the second for th	ogram to retrie covered by a to de for general nd usage. For rough annotati dits page for th equence and A ummal genes and Gene P wnGene genome genome s (names/acce tate cn: create m: create	eve the data asso rack. For help in u information and s more complex que on enrichments, s e list of contributo Annotation Downle genome: Human redictions ENCODE Pilot reg essions): paste list from selected table	ciated with a trac sing this applica ample queries, a eries, you may w end the data to rs and usage res bads page. • asse track: UCSC Ger describe table scher gions position at upload list	k in text format, to tion see <u>Using the</u> and the OpenHelix T ant to use <u>Galaxy of</u> <u>GREAT</u> . Send data strictions associated embly: Feb. 2009 (Gl nes Constructions n chr14:54416455-544	Calculate inter Table Browser Table Browser or our <u>public M</u> to <u>GenomeSp</u> d with these da RCh37/hg19) • add custom track 23554 lookup <u>Galaxy</u> <u>Galaxy</u>	sections b t <u>utorial</u> for <u>tySQL ser</u> <u>bace</u> for us ata. All tab track h define r	etween tracks, an cription of the con- r a narrated prese ver. To examine the with diverse con- oles can be downlo ubs regions	nd to retrieve DNA trols in this form, th entation of the softw ne biological functio mputational tools. F boaded in their entire	e vare on of Refer ety	
output file	e:	all fields	from selected	table	·					
file type r	eturned: • p	lain te selected	fields from pri	mary and related	l tables					
get output To reset al	summary/stati	istics GTF - ge ings (i Custom t	e ne transfer for owser extensib rack	mat ole data						
		hyperlink	s to Genome I	Browser						

Demo and exercise 3

- Load peaks (bed format) derived from ChIPseq:
 - <u>GM12878 H3K36me3 Histone Mods by ChIP-seq Peaks</u> <u>from ENCODE/Broad</u>
 - To save time, only peaks in chr22
- Identify the Refseq genes that could be regulated by H3K36me3
 - Go to the Table Browser:
 - Choose RefGene table
 - Intersect with the above uploaded track



Tools: LiftOver

Genome/Build Conversion



LiftOver: Genome/Build Conversion – Multiple Regions







Tools: blat

21 463 483 1015 100.0%

20 903 922 1015 100.0%



browser details my_seq

browser details my_seq

WHITEHEAD INSTITUTE

Ba

SPAN

4 + 188270839 188270859

1 - 29594014 29594033

Tools: Gene Sorter*

bv criteria: e	eg. similarity.	Enter gene	Enter gene of interest		
e expression					
Genomes Genome Browser Tools JCSC Human Gene Sorter	Mirrors Downloads My Date Help , genome Help , soft by Frotein Homology	About Us assembly Feb. 2009 (GRCh37/hg19) BLASTP - configure filter (now off	• search display 50 • output sequence text		
About the Gene Sorter					
This program displays a sorted table of genes tha	at are related to one another. The relationship can be o	one of several types, including protein	-level homology, similarity of gene expression profiles, or genomic proximity,		
To display a gene and its relatives:		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,			
 Select a genome and assembly from the cordination of phrase into the search text bits. Choose the gene relationship with which you 4. Press the Gol button to display your results. 	irresponding pull-down menus. box to specify which gene should be displayed in the G ou would like to sort the list by selecting an option from .	Sene Sorter. Examples of search terr the <i>sort by</i> pull-down menu.	s include FOXA2, HOXA9, and MAP kinase.		
Following a successful search, the Gene Sorter d	lisplays a table containing the specified gene highlig	hted in light green and its relatives	each on a separate line. To adjust the number of rows shown, select an option from the display pull-down	menu.	
The default set of table columns which can be e to the highlighted gene. To select a different gene information about the gene.	expanded, reduced, and rearranged via the <i>configure</i> t a in the list, click on its name. Clicking on a gene's Gen	button shows additional information nome Position will open the UCSC G	about the genes. Some of the column data, such as those in the <i>BLASTE-value</i> and %/ <i>D</i> columns, are ce nome Browser to the location of that gene. Similarly, clicking on a gene's <i>Description</i> will open a page short	alculated relative wing detailed	
One of the most powerful features of the Gene So to select all human genes over-expressed in the o	orter is its filtering capabilities, accessed via the <i>filter</i> b cerebellum that have GO-annotated G-protein coupled	utton. Use the filter to fine-tune the li I receptor activity.	t of displayed genes to a subset based on a selection of detailed and flexible criteria. For example, the filte	er may be used	
The Gene Sorter offers two options for displaying	and downloading sequence associated with the gener	s in the table. Clicking on the sequer	ce button will fetch associated protein, mRNA, promoter, or genomic sequence. To dump the table into a si	mple	
The UCSC Gene Sorter was designed and impler Howard Hughes Medical Institute.	mented by Jim Kent, Fan Hsu, Donna Karolchik, David	I Haussler, and the UCSC Genome E	ioinformatics Group. This work is supported by a grant from the National Human Genome Research Institu	te and by the	
ī	UCSC Human Gene Sorter				
			genome Human • assembly Dec. 2013 (GRCh38/hg38) • Search uc064wua.1	Go!	
			sort by Protein Homology - BLASTP Configure filter (now off) display 25 Output sequ	ence text	
	Numecy Vision	er BLASTP Versty	on Description		
ļ	1ABO n/a	0 chr9 133,265,	108 The sequence shown here is derived from an Ensembl automatic analysis pipeline and s	hould be considered as preliminary da	
	2 GBGT1 n/a n/a n/a n/a n/a	4e-70 chr9 133,158,	149 Homo sapiens globoside alpha-1,3-N-acetylgalactosaminyltransferase 1 (GBGT1), transc	ript variant 1, mRNA. (from RefSeq NI	
		10-52 CHIT 33,313,9	22 Inonio sapiens aipira 1,3-galaciosylitaristerase 2 (A3GALT2), mikiva. (from RefSeq NM 1	- NM 402074)	



Prot A0A087X009)

Track Data Hubs

â	Genomes	Genome Browser	Tools	Mirrors	Downloads	My Data	Help	About Us	
Track Da	ata Hubs					Sessions			
						Track Hub	5		
Track da	ta hubs are c	ollections of externa	l tracks th	at can be im	ported into the	Custom In	acks	rack	s show up under the hu
main bro	wser page, a	s well as on the cont	igure pag	e. For more	information, see	e the User's Q	Suide. Io I	import a publi	c hub click its "Connect

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

Public Hubs	My Hubs								
Enter search terms to find in public track hub description pages:									
methylation Search Public Hubs									
Displayed list restricted by search terms: methylation Show All Hubs									
Display	Display Hub Name Description Assemblies								
	Deadman Enicenemics Data	Poadman Enigonomics Human Enigonomo Atlas							
Connect	Complete Collection at Wash U VizHub	Data Complete Collection, VizHub at Washington University in St. Louis	hg19						
Connect	Complete Collection at Wash U VizHub ENCODE Analysis Hub	Data Complete Collection, VizHub at Washington University in St. Louis ENCODE Integrative Analysis Data Hub	hg19 hg19						
Connect Connect Connect	Complete Collection at Wash U VizHub ENCODE Analysis Hub DNA Methylation	Noadmap Epigenomics Human Epigenomic Atlas Data Complete Collection, VizHub at Washington University in St. Louis ENCODE Integrative Analysis Data Hub Hundreds of analyzed methylomes from bisulfite sequencing data	hg19 hg19 [+] hg38, hg19, hg18, mm9, mm10, panTro2						



Share/Save Session

Â	Genomes	Genome Browser	Tools	Mirrors	Downloads	My Data	Help	About Us	
Sign in	to UCSC Ge	nome Bioinforma	atics			Sessions			
Login						Track Hubs			
Create a	an account					Custom Tra	icks		
Signing	in enables voi	i to save current set	tinas into a	a named se	ession and then	restore settin	as from th	e session later	
If you wi	ish, you can sl	hare named session	s with othe	er users.			go nom a		
Sessior	n Manageme	ent							
See the	Sessions Use	e <mark>r's Guide</mark> for more ir	nformation	about this	tool.				
Click he	<u>Click here to reset</u> the browser user interface settings to their defaults.								
lf you <u>si</u> g	<u>gn in</u> , you will	also have the option	to save na	amed sessi	ions.				
Save S	Settings								
Save cu	irrent settings	to a local file:							
file:		file type retur	ned: plain	text	- submit				
(leave	e file blank to	get output in browse	r window)						
Restor	e Settings								
Use set	tings from and	other user's saved se	ession:						
user:		session na	me:		submit				
Use set	tings from a lo	ocal file:	Browse	No file sel	ected.	submit			
Use set	tings from a U	RL (http://, ftp://)			submit				

Ensembl ensembl.org

- Website was launched in 2000
- Joint project by EMBL-EBI and Wellcome Trust
- Current Ensembl release: 89 (May 2017)
- Not just a browser an automated genome annotation pipeline
 - includes hand-curated gene annotation (eg. Vega and Havana)
- BioMart allows easy access to (download) most of the Ensembl data





Ensembl Browser



Ensembl: Gene Tree

Location: 9:133,250,401-133,275,214









BioMart

CENSEMBI BLAST/BLAT	BioMart Tools Downloads Help & Documentation Blog	Login/Register द्वी ← Search all species Q
New Count Results	∳ URL S XML S Peri ® Help	
Dataset Human genes (GRCh38,p10) Filters [None selected] Attributes Gene stable ID Transcript stable ID Dataset [None Selected]	Choose Database: • Ensembl Genes • Mouse Strains • Ensembl Variatio • Ensembl Regulat	n ion

_

Whitehead Institute



BioMart: Filter for Input

	AT BioMart Tools Downloads Help & Documentation Blog		Login/Register
New Count Results		🖕 URL 🛛 XML 🖉 Peri 🕲 Help	
Dataset Human geoss (GRCh38.p10)	(If filter value	Please restrict your query using criteria below use are truncated in any lists, hover over the list item to see the full text)	
[None selected]			
Gene stable ID Transcript stable ID	Limit to genes (external references)	With CCDS ID(s)	
Dataset [None Selected]	Input external references ID list [Max 500 advised]	Gene stable ID(s) [e.g. ENSC0000000003] ▼ Gene stable ID(s) [e.g. ENSC0000000003] ▲ Transcript stable ID(s) [e.g. ENST00000000233] ▲ Protein stable ID(s) [e.g. ENSP0000000233] ■ Exon ID(s) [e.g. ENSE00000327880] ■	
Filter by	Limit to genes (microarray probes/probesets)	Gene Name(s) [e.g. sno26] Transcript Name(s) [e.g. sno26.3-201] CCDS ID(s) [e.g. CCDS10] ded	
 Region 	Input microarray probes/probesets ID list [Max 500 advised]	ChEMBL ID(s) [e.g. CHEMBL1075092] Clone-based (Ensembl) gene ID(s) [e.g. AB015752.1] Clone-based (Ensembl) transmit ID(s) [e.g. AB015752.1-201]	
• Gene		Clone-based (Vega) gene ID(s) [e.g. AB015752.3] Clone-based (Vega) transcript ID(s) [e.g. AB015752.3-001]	
 Phenotype 	Transcript count >=	DataBase of Aberrant 3' Spice Sites name(s) [e.g. ABCC8] DataBase of Aberrant 3' Spice Sites ID(s) [e.g. 1] DataBase of Aberrant 5' Spice Sites name(s) [e.g. ABCD1]	
• GO	Transcript count <= Gene type	Data8ase of Aberrant 5' Spice Stes ID(s) [e.g. 1] Ensembl Human Transcript ID(s) [e.g. EHST0000003084] Ensembl Human Transcript ID(s) [e.g. EHST00000003084]	
Multi species	S	EntrezGene transcript name ID(s) [e.g. AA06-201] European Nucleotide Archive ID(s) [e.g. AA06-201]	
comparison		IG_D_gene IG_J_gene IG_J_pseudogene IG_pseudogene	
Protein	Transcript type	3prime_overlapping_ncRNA	
domains/fan	n	IG C gene	

• Variant



BioMart: Attributes for Output

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

🛃 🕶 Search all species...

Q

New Count Results		★ URL 2 Perl ® Help	
Dataset		Please select columns to be included in the output and hit 'Results' when ready	
Human genes (GRCh38.p10)	Missing non coding genes in your mart query output, please check the following FAQ		
Filters [None selected] Attributes	✔ Features ● Variant (Germline) Structures ● Variant (Somatic) ● Homologues ● Sequences		
Gene stable ID Transcript stable ID	GENE:		
	Ensembl	APPRIS annotation	
Dataset	Transcript stable ID	Gene name	
[None Selected]	Protein stable ID Exon stable ID Gene description Gene start (bp) Gene end (bp) Strand Karyotype band Transcript start (bp) Transcript end (bp) Transcript end (bp) Transcript ength (including UTRs and CDS) Transcript support level (TSL) GENCODE basic annotation	Source of gene name Transcript name Source of transcript name Transcript count % GC content Gene type Transcript type Source (gene) Source (transcript) Status (grene) Status (transcript) Version (gene) Version (transcript)	
	Phenotype Phenotype description Source name Study external reference EXTERNAL: B PROTEIN DOMAINS AND FAMILIES:	 Strain name Strain gender □ P value 	

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

DiaMart futorial | VouTuba | VouKu





BioMart: Attributes for Output

CENSEMBI BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog

Login/Register

New Count Results	🛧 URL 😰 XML 🔮 Perl 🕐 Help
Dataset	Please select columns to be included in the output and hit 'Results' when ready
Human genes (GRCh38.p10) Filters Phone selected] Attributer Peptide Gene stable ID Transcript stable ID	Missing non coding genes in your mart query output, please check the following FAQ
	Features Variant (Germline) Structures Variant (Somatic) Homologues Sequences
	B SEQUENCES:
	Sequences (max 1)
Dataset	
[None Selected]	Unspliced (Transcript) 5' UTR Unspliced (Gene) 3' UTR Flank (Transcript) Exon sequences Flank (Gene) cDNA sequences Flank-coding region (Transcript) Coding sequence Flank-coding region (Gene) @ Peptide
	Upstream flank
	Downstream flank
	HEADER INFORMATION:

Datasets -> Filters (filtering and inputs) -> Attributes (desired output) -> Results

RioMart futorial | YouTube | YouKu



1



Practical Tips

- Take careful notes of genome assembly (eg. hg19 vs hg38) for
 - coordinates
 - custom browser files
- Genome assembly are usually updated every several years, however, annotation may be updated frequently (eg. RefSeq)
- Data displayed, especially on UCSC Genome Browser, may be generated from other sources (click on track description for more info or citation)
- Try different genome browsers based on your questions and needs
 - turning-on all available tracks/data will be too noisy: use relevant tracks only



More Information

UCSC Browser Tutorials

https://genome.ucsc.edu/training/index.html

Free Videos:

https://genome.ucsc.edu/training/vids/index.html

Open Helix:

http://www.openhelix.com/ucsc

- FAQ: https://genome.ucsc.edu/FAQ/
- Genomewiki:

http://genomewiki.ucsc.edu/index.php/Main_Page

- Kent Utils (command-line tools already installed on tak): http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/FOOTER
- Ensembl Tutorials:

http://www.ensembl.org/info/website/tutorials/index.html



