

Identifying potential candidate genes

1. Make sure that you correctly carried on the marker name information to your FST calculation
 - a. Do this by ordering your map file and genotype file the same way and then making the rownames of your genotype file, your FST matrix, etc. the SNP name of your map file
 - b. `Rownames(geno)=map$snp`
2. Once you've got your FST peaks and regions of interest (both in GWAS and diversity), look up the rownames (the name of the SNP that is of interest), and refer back to your map file where you will find the chromosome and exact base pair position that the SNP is located at.
3. Take the position information across to Ensembl genome browser <https://asia.ensembl.org/index.html>
 - a. As a rule of thumb, look 1,000,000 base pairs up and down stream of your marker. Let's say your marker is at position 3,000,000, then look for genes between 2,000,000-4,000,000 base pairs

The screenshot shows the Ensembl genome browser homepage. The browser's address bar displays `https://asia.ensembl.org/index.html`. The top navigation bar includes links for `BLAST/BLAT`, `VEP`, `Tools`, `BioMart`, `Downloads`, `Help & Docs`, and `Blog`.

Below the navigation bar, there are several sections:

- Tools**: Includes a link to `All tools`.
- BioMart**: Described as "Export custom datasets from Ensembl with this data-mining tool".
- BLAST/BLAT**: Described as "Search our genomes for your sequence".
- Variant Effect Predictor**: Described as "Predict the effect of known and predicted variants".

The main search area is highlighted with a red box and contains the following text:

Search
Cow for
2:6270000-6280000
Go
e.g. `BRCA2` or `rat 5:62797383-6367669` or `rs699` or `coronary heart disease`

Two red arrows point from orange text boxes to the search area:

- One arrow points to the "Cow" dropdown menu, with the text: "Look for 'cow', as this is the species you are working with".
- Another arrow points to the "2:6270000-6280000" text input, with the text: "Put in your position you are interested in."

Below the search area, there is a section for "All genomes" with a dropdown menu set to "-- Select a species --". Below this are links to "View full list of all Ensembl species" and "Edit your favourites".

To the right of the "All genomes" section is a "Favourite genomes" section with three entries:

- Human (GRCh38)
- Mouse (GRCm38.p6)
- Zebrafish (GRCz11)

On the far right, there is a section titled "Ensembl Release 96" with a list of updates and a section titled "Other news from our" with a list of recent news items.

Primary_assembly 2: 6,270,000-6,280,000

Location: 2:6,270,000-6,280,000

Region in detail

Location: 2:6270000-6280000

Gene: MSTN

Gene Legend

Ensembl protein coding

88 way GERP elements

Proteins from UniProtKB

cDNA alignment

EST cluster (Unigene)

Genes (Ensembl)

Contigs

All sequence SNPs/...

SV - Smaller variants

%GC

Gene Legend

Protein Coding

Motif Feature Legend

Variant Legend

There are currently 62 tracks turned off.

Ensembl Bos taurus version 95.12 (ARS-UCD1.2) Primary_assembly 2: 6,270,000 - 6,280,000

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Policy](#) and [Terms of Use](#) I Agree

Red blocks are known protein coding genes, the red horizontal box indicates the start and end position that you specified.

1. As you can see, there is a gene called MSTN touching the region that we are interested in (this is not based on your project, and I specifically picked a region where I knew that there would be an interesting gene).
 - a. There are more interesting features on this page, for example in the box underneath, known SNPs for this regions are indicated, and we can see that there are some missense and synonymous variant known to be located in the MSTN gene.
 - b. You can also move left and right along the gene sequence or zoom in and out to get a more global or precise look at your region of interest.
2. You can click on the gene name to get more information about it.
 - a. Its full name is myostatin, you see its Ensembl gene id, and its exact start and end positions
 - b. Click on the Ensembl gene Id and it will give you more information:
 - i. It has 3 exons
 - ii. 1 splice variant
 - iii. 128 orthologues
 - iv. 32 paralogues
 - v. Belongs to 1 protein family
 - vi. Is associated with 1 phenotype

Gene: MSTN (ENSBTAG00000011808)

Description: myostatin [Source:VGNC Symbol;Acc:VGNC:31709]

Location: Primary assembly 2: 6,278,630-6,285,486 forward strand. ARS-UCD1.2:CM008169.2

About this gene: This gene has 1 transcript (splice variant), 128 orthologues, 32 paralogues, is a member of 1 Ensembl protein family and is associated with 1 phenotype.

Summary

Name: MSTN (VGNC Symbol)

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: Q18836

Ensembl version: ENSBTAG00000011808.5

Gene type: Protein coding

Annotation method: Annotation produced by the Ensembl genebuild.

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Genes (Ensembl): 6.270Mb 6.275Mb 6.280Mb 6.285Mb 6.290Mb

Contigs: 6.270Mb 6.275Mb 6.280Mb 6.285Mb 6.290Mb

Gene Legend: 26.86 kb

MSTN-201 > protein coding

Reverse strand

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Policy](#) and [Terms of Use](#) I Agree

c. Click on phenotype

- The trait is muscular hypertrophy causing double muscling (Belgian Blue cattle, but also other mammals)
- Further down, you see other traits that orthologues of the MSTN gene are associated with

You can get more information on the gene by clicking your way through the phenotypes.

Another way is to take the gene name MSTN across to NCBI <https://www.ncbi.nlm.nih.gov/pubmed/> (leaving it at 'pubmed', this is a great data base for scientific publications for your introduction and discussion)

Home - PubMed - NCB

<https://www.ncbi.nlm.nih.gov/pubmed/>

ONLINE - AK Google Home - PubMed - NC English = Deutsch V Assessing dairy cows Textbook - Animal Br KEGG - Table of Cont Ensembl Genome Br Snp Chimp introductiontoparent session - SmartFees

NCBI Resources How To Sign in to NCBI

PubMed.gov US National Library of Medicine National Institutes of Health

Gene MSTN Advanced Search

Switch to 'Gene', type in your gene name

PubMed comprises more than 29 million citations for biomedical literature from MEDLINE, life science journals, and online Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

PubMed Quick Start Guide

Full Text Articles

PubMed FAQs

PubMed Tutorials

New and Noteworthy

PubMed Tools

PubMed Mobile

Single Citation Matcher

Batch Citation Matcher

Clinical Queries

Topic-Specific Queries

More Resources

MeSH Database

Journals in NCBI Databases

Clinical Trials

E-Utilities (API)

LinkOut

Latest Literature

New articles from highly accessed journals

Trending Articles

PubMed records with recent increases in activity

https://www.ncbi.nlm.nih.gov/gene/?term=MSTN

E-Mail-Adn SPIEGEL ONLINE - Ak Google Home - PubMed - NC Englisch = Deutsch V Assessing dairy cows Textbook - Animal Br KEGG - Table of Cont Ensembl Genome Brc Snp Chimp introductio

See also 5 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> Mstn ID: 17700	myostatin [<i>Mus musculus</i> (house mouse)]	Chromosome 1, NC_000067.6 (53061640..53068079)	Cmpt, Gdf8	
<input type="checkbox"/> MSTN ID: 2660	myostatin [<i>Homo sapiens</i> (human)]	Chromosome 2, NC_000002.12 (190055700..190062729, complement)	GDF8, MSLHP	601788
<input type="checkbox"/> Mstn ID: 29152	myostatin [<i>Rattus norvegicus</i> (Norway rat)]	Chromosome 9, NC_005108.4 (53310977..53315804, complement)	Gdf8	
<input type="checkbox"/> MSTN ID: 281187	myostatin [<i>Bos taurus</i> (cattle)]	Chromosome 2, NC_037329.1 (6278864..6285491)	GDF8	
<input type="checkbox"/> MSTN ID: 443449	myostatin [<i>Ovis aries</i> (sheep)]	Chromosome 12, NC_007093.1 (120000000..120000000, complement)	GDF8, GDF8	
<input type="checkbox"/> MSTN ID: 373964	myostatin [<i>Gallus gallus</i> (chicken)]	Chromosome 1, NC_000963.3 (100000000..100000000, complement)	GDF8, GDF8	
<input type="checkbox"/> MSTN ID: 399534	myostatin [<i>Sus scrofa</i> (pig)]	Chromosome 2, NC_007093.1 (120000000..120000000, complement)	GDF8, GDF8:MSTN	
<input type="checkbox"/> MSTN ID: 100033832	myostatin [<i>Equus caballus</i> (horse)]	Chromosome 18, NC_009161.3 (66605150..66610122, complement)		
<input type="checkbox"/> MSTN ID: 100860887	myostatin [<i>Capra hircus</i> (goat)]	Chromosome 2, NC_030809.1 (130227819..130232923, complement)	GDF-8, GDF8	
<input type="checkbox"/> MSTN ID: 100125998	myostatin [<i>Oryctolagus cuniculus</i> (rabbit)]	Chromosome 7, NC_013675.1 (130429151..130434062, complement)		
<input type="checkbox"/> MSTN ID: 403433	myostatin [<i>Canis lupus familiaris</i> (dog)]	Chromosome 37, NC_006619.3 (729172..734255, complement)	GDF8	
<input type="checkbox"/> MSTN ID: 101800742	myostatin [<i>Anas platyrhynchos</i> (mallard)]	Chromosome 7, NC_040052.1 (8408607..8416472)	Anapl_10019, GDF-8, GDF8	
<input type="checkbox"/> mstn ID: 100529165	myostatin [<i>Oryzias latipes</i> (Japanese medaka)]	Chromosome 21, NC_019879.2 (16518077..16520334)		
<input type="checkbox"/> MSTN ID: 102414042	myostatin [<i>Bubalus bubalis</i> (water buffalo)]	Chromosome 2, NC_037546.1 (58758414..58763512)	GDF-8, GDF8	
<input type="checkbox"/> MSTN ID: 710114	myostatin [<i>Macaca mulatta</i> (Rhesus monkey)]	Chromosome 12, NC_041765.1 (77245473..77250906, complement)	GDF8	
<input type="checkbox"/> MSTN ID: 470605	myostatin [<i>Pan troglodytes</i> (chimpanzee)]	Chromosome 2B, NC_036881.1 (76667661..76672993, complement)	CK820_G0053701, GDF8	

Make sure you select the gene for the species that you are working in (Bos Taurus)

1. Again, you will find a lot of information about the gene, its position and structure
2. Scroll down and you find a tab called "Bibliography" which will list all scientific articles that mention the gene (if there are any)
3. Further down, you'll find a tab "Pathways from BioSystems" which gives you information about the genetic network that the gene is involved in and how it exerts its function
4. Underneath that, under "General gene information", you will find "Gene Ontology", which provides information about the biological function of the gene, biological processes that the gene is involved in, and what the encoded protein is part of (Component)
5. Again, you can click your way through to more information as you need it.