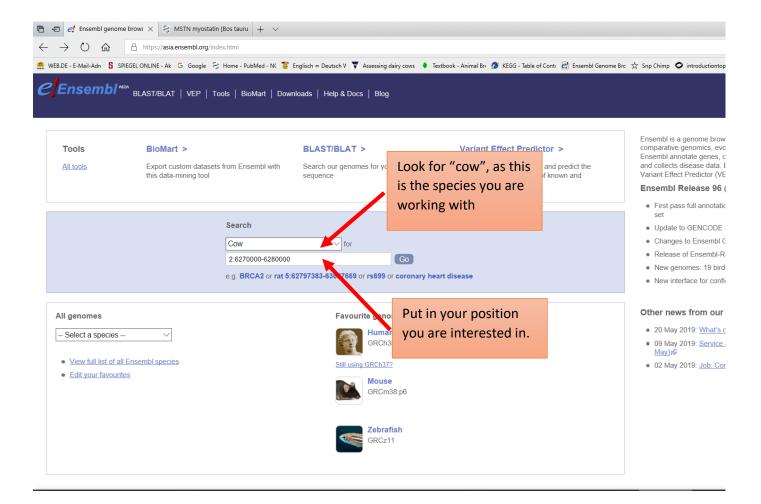
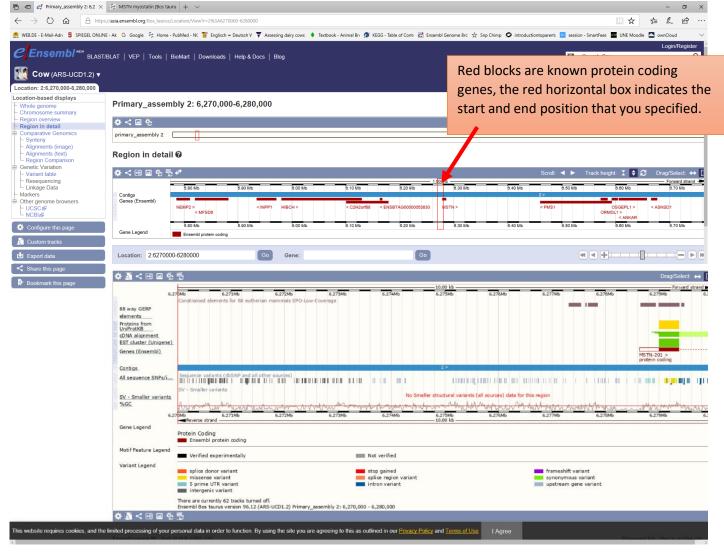
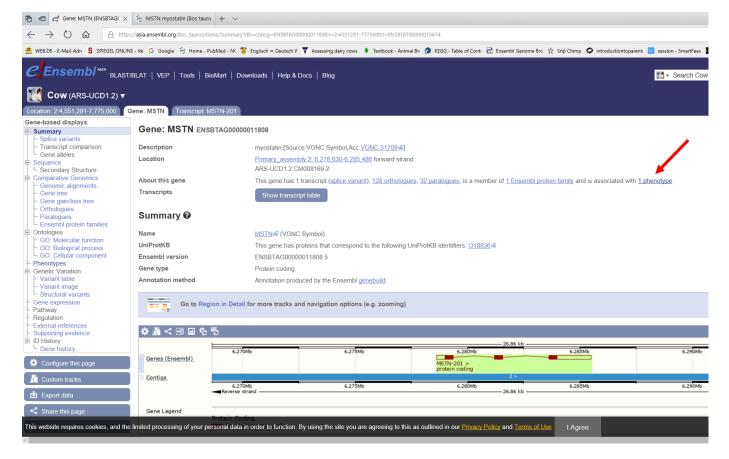
Identifying potential candidate genes

- 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





- 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

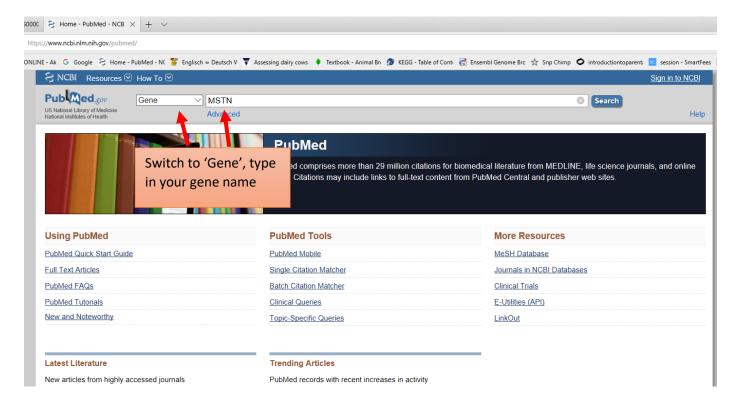


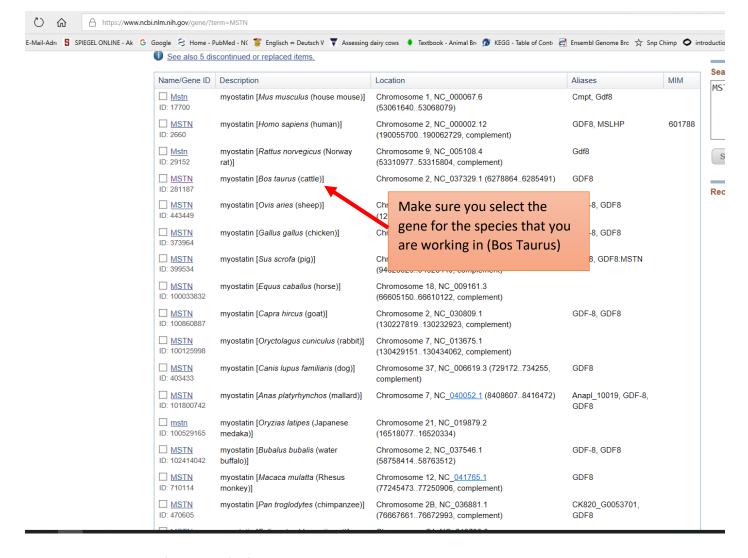
c. Click on phenotype

- i. The trait is muscular hypertrophy causing double muscling (Belgian Blue cattle, but also other mammals)
- ii. 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)





- 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.