Xiaoxu Na

**Genomic regions, genes and transcripts**

1. What is the name of the genome assembly for Panda? How long is the Panda genome? How many genes have been annotated?

ailMel1 (GCA\_000004335.1)

2,299,509,015  
19,343

1. Search for human BRCA2 gene.

* Create a Share link for this display and email the link to. me and open the link.
* Download genomic sequence of BRCA2 in FASTA format.

<http://useast.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000139618;r=13:32315474-32400266>

1. (a) Find the human MYH9 gene.

>On which chromosome and which strand of the genome is this gene located [Chromosome 22: 36,281,277-36,388,067](http://useast.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG00000100345;r=22:36281277-36388067) reverse strand.

> How many transcripts are there?

This gene has 11 transcripts.

>What is the longest transcript, and how long is the protein it encodes?

MYH9-201, 7554bp

>Which transcript has a CCDS record associate with?

MYH9-201

(b) Are there any diseases associated with this gene, according to MIM?

Association with Kidney Disease in African Americans

(c) In the transcript table, choose the transcript ID for MYH9-201 and go to the Transcript tab.

> How many exons does it have?

This transcript has [41 exons](http://useast.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000100345;r=22:36281277-36388067;t=ENST00000216181)

> Is there an associated sequence in UniProtKB/Swiss-Prot?

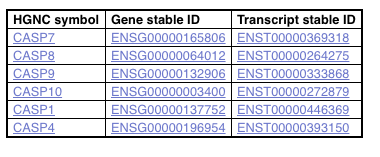
Yes, UniProtKB - P35579 (MYH9\_HUMAN)

>What are some functions for MYH9-201 according to the Gene Ontology consortium’

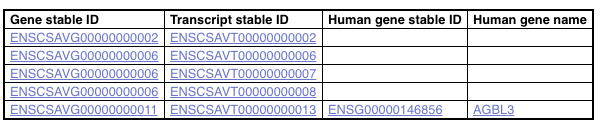
microfilament motor activity, nucleotide binding, RNA binding

**BioMart**

1. Give a list of 6 IDs of human proteins from the NCBI RefSeq database: NP\_001218, NP\_150636, NP\_150649, NP\_001220, NP\_116756, NP\_001219, generate a list that shows to which Ensembl HumanTanscript IDs and to which HGNC symbols these RefSeq IDs correspond.



1. For a list Ciona savignyi Ensembl genes, ENSCSAVG00000000002, ENSCSAVG00000000006, ENSCSAVG00000000011, export the human orthologues



**Variants**

1. The SNP rs1738074 in the 5'UTR of the human TAGAP gene has been identified as a genetic risk factor for a few diseases.
2. In which transcripts is this SNP found?

[ENST00000326965.7](http://uswest.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000164691;source=dbSNP;t=ENST00000326965;v=rs1738074;vdb=variation;vf=104056402) , [ENST00000338313.5](http://uswest.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000164691;source=dbSNP;t=ENST00000338313;v=rs1738074;vdb=variation;vf=104056402) , [ENST00000367066.7](http://uswest.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000164691;source=dbSNP;t=ENST00000367066;v=rs1738074;vdb=variation;vf=104056402), [ENST00000642909.1](http://uswest.ensembl.org/Homo_sapiens/Transcript/Variation_Transcript/Table?db=core;g=ENSG00000164691;source=dbSNP;t=ENST00000642909;v=rs1738074;vdb=variation;vf=104056402)

1. What is the least frequent genotype for this SNP in the Yoruba(YRI) population from the HapMap set?

C|C: 0.097 (11)

1. What is the ancestral allele? Is it conserved in the 12 primates EPO alignment? Is it conserved in the 26 eutherian mammals EPO?

T: 0.695 (157)

C: 0.305 (69)

Yes, Yes

1. With which diseases is this SNP associated?

Celiac disease, multiple sclerosis